

UNIVERSIDADE FEDERAL DE SANTA CATARINA

TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

**CÉSAR HENRIQUE CARDOSO DOS SANTOS**

**MAYKON CARLOS DE FREITAS**

**VISDADOSENEM: VISUALIZAÇÃO DE DADOS DO ENEM COM A  
BIBLIOTECA D3**

**Araranguá, 08 de dezembro de 2016**

CÉSAR HENRIQUE CARDOSO DOS SANTOS  
MAYKON CARLOS DE FREITAS

## **VISDADOSENEM: VISUALIZAÇÃO DE DADOS DO ENEM COM A BIBLIOTECA D3**

Trabalho de Curso submetido à Universidade Federal de Santa Catarina como parte dos requisitos necessários para a obtenção do Grau de Bacharel em Tecnologias da Informação e Comunicação. Sob a orientação do Professor Robson Rodrigues Lemos.

**Araranguá, 2016**

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Santos, César

VisDadosEnem : Visualização de dados do ENEM com a  
biblioteca D3 / César Santos ; orientador, Robson Rodrigues  
Lemos - Araranguá, SC, 2016.

73 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Campus Araranguá.  
Graduação em Tecnologias da Informação e Comunicação.

Inclui referências

1. Tecnologias da Informação e Comunicação. 2.  
visualização de dados. I. Lemos Rodrigues, Robson. II.  
Universidade Federal de Santa Catarina. Graduação em  
Tecnologias da Informação e Comunicação. III. Título.

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

de Freitas, Maykon Carlos

VisDadosEnem : Visualização de dados do ENEM com a  
biblioteca D3 / Maykon Carlos de Freitas ; orientador,  
Robson Rodrigues Lemos - Araranguá, SC, 2016.

73 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Campus Araranguá.  
Graduação em Tecnologias da Informação e Comunicação.

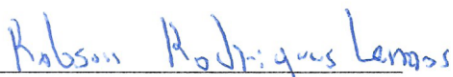
Inclui referências

1. Tecnologias da Informação e Comunicação. 2.  
Visualização de dados. I. Rodrigues Lemos, Robson. II.  
Universidade Federal de Santa Catarina. Graduação em  
Tecnologias da Informação e Comunicação. III. Título.

CÉSAR HENRIQUE CARDOSO DOS SANTOS  
MAYKON CARLOS DE FREITAS

**VISDADOSENEM: VISUALIZAÇÃO DE DADOS DO ENEM COM A  
BIBLIOTECA D3**

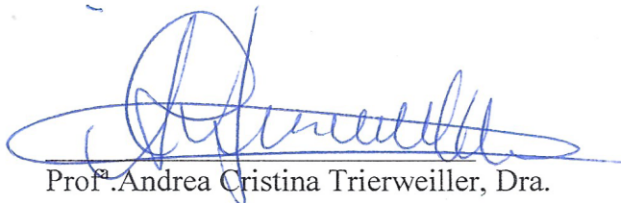
Trabalho de Curso submetido à Universidade Federal de Santa Catarina, como parte dos requisitos necessários para a obtenção do Grau de Bacharel em Tecnologias da Informação e Comunicação.



Prof. Robson Rodrigues Lemos, Dr.  
Orientador  
Universidade Federal de Santa Catarina



Prof. Alexandre Leopoldo Gonçalves, Dr.  
Revisor  
Universidade Federal de Santa Catarina



Prof.ª Andrea Cristina Trierweiler, Dra.  
Revisora  
Universidade Federal de Santa Catarina

## AGRADECIMENTOS

*Agradeço aos meus pais por todo apoio que sempre me deram. Ao professor Robson Rodrigues Lemos que sem dúvidas teve uma participação fundamental na elaboração deste trabalho. Ao meu colega Maykon Carlos de Freitas por todo seu empenho e dedicação. E a todos os meus familiares, amigos e colegas.*

*César Henrique Cardoso dos Santos*

*Primeiramente quero agradecer a minha família, em especial aos meus pais, onde através deles recebi todo apoio necessário para a realização do curso. A todos meus colegas e professores que ao longo dessa caminhada, proporcionaram um excelente ambiente de aprendizagem. Sem esquecer é claro, do professor Robson Lemos Rodrigues que teve uma participação fundamental orientando este trabalho, assim como meu colega de TCC César Henrique pelo tamanho empenho e dedicação.*

*Maykon Carlos de Freitas*

*O sucesso nasce do querer, da determinação e  
persistência em se chegar a um objetivo.  
Mesmo não atingindo o alvo, quem busca e  
vence obstáculos, no mínimo fará coisas  
admiráveis.*

***José de Alencar***

## RESUMO

Com o avanço tecnológico surgem soluções capazes de resolver problemas com alto nível de complexidade, grande volume de dados e que demandam um alto poder de processamento. O aumento em massa da informação pode gerar alguns problemas para análise e compreensão de dados. Os computadores evoluem em grande escala, já os usuários não possuem a capacidade de acompanhar tamanha evolução. Pensando nisso, foi elaborado este trabalho com o objetivo de demonstrar a importância da visualização da informação sobre um grande volume de dados. Para tal, foi desenvolvida uma aplicação Web para visualização de dados do exame nacional do ensino médio (ENEM) com a biblioteca D3 (*Data-Driven Documents*), na qual seja possível disponibilizá-la na internet, para facilitar o processo de análise dos dados e tomada de decisão do exame nacional do ensino médio. Levando em consideração os recursos de visualização que a aplicação proporcionou, comprovou-se que o seu uso, ou o uso de ferramentas similares, faz-se necessário para realizar análises sobre um grande conjunto de dados, como no caso dos dados adotados neste trabalho. Análise esta, que uma tabela comum dificilmente irá proporcionar. O estudo das tecnologias da informação e comunicação possibilitou o desenvolvimento da ferramenta para visualização de dados, intitulado VisDadosEnem. Como resultado das análises comparativas realizadas sobre o ENEM com o VisDadosEnem, observou-se que alunos de escolas particulares, ou que cursaram maior parte do tempo nela, obtiveram no geral uma melhor média em comparação à alunos oriundos de escola pública. O projeto foi desenvolvido com os micro dados do ENEM, a partir do portal brasileiro de dados abertos, para comprovar a necessidade e eficácia da utilização de ferramentas nesse sentido. Pode-se afirmar que, a utilização do VisDadosEnem possibilitou uma análise comparativa sobre o conjunto de dados e foram descobertos padrões que seriam dificilmente encontrados em uma tabela ou gráfico tradicional.

Palavras-chave: Visualização da Informação, Análise de Dados, Micro Dados, Tecnologias da Informação e Comunicação e Dados do ENEM.



## **ABSTRACT**

With the technological advances, they are solutions capable of solving problems with high level of complexity, great volume of data and that demand a great volume of processing. The increase in the mass of the information can generate some problems for analysis and understanding of data. Computers evolve on a large scale, and users do not have the capacity to keep pace with such evolution. Based on that, the main purpose of this study is to investigate the importance of visualizing the information regarding a large volume of data. In order to do that, a Web application was developed for the visualization of national exam of the high school (ENEM) data with the D3 (Data-Driven Documents) library. The Web application facilitate the data analysis and decision making process for the national exam of the high school. Taking into account the visualization capabilities provided by the application, it has been proven that its use, or the use of similar tools, it is necessary to perform analyzes on a large data set, as well as in the case of the data adopted in this study. For that type of analysis making use of a traditional table would be difficult to obtain similar results. The study of information and communication technologies enabled the development of the data visualization tool, entitled VisDadosEnem. As a result of the comparative analyzes carried out on the ENEM with VisDadosEnem, it was observed that students from private schools, or who spent most of their time in that particular type of school, generally obtained a better average score compared to other schools. The project was developed with the ENEM micro data from the Brazilian open data portal, in order to prove the need and effectiveness of the use of visualization information tools. It can be stated that the use of VisDadosEnem enabled a comparative analysis on the data set and the ability to visualize patterns that would be difficult to identify in a traditional table or chart.

**Keywords:** Information Visualization, Data Analysis, Micro Data, Information and Communication Technologies and ENEM Data.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Técnica de visualização de dados baseada em pontos.....	23
Figura 2 - Técnica de visualização de dados baseada em linhas.....	23
Figura 3 - Técnica de visualização de dados baseada em regiões.....	24
Figura 4 - Combinações de técnicas sobre dados multivariados.....	25
Figura 5 - Método de preenchimento de espaço.....	26
Figura 6 - Método de não preenchimento de espaço.....	26
Figura 7 - Técnica de redes e grafos com restrições baseadas em forças. ....	27
Figura 8 - Técnica de grafos e redes baseada em atributos. ....	28
Figura 9 - Visualização da recuperação da informação. ....	29
Figura 10 - Visualização de coleção de documentos. ....	29
Figura 11 – Etapas do processo de visualização baseada no modelo de referência de Card (1999).....	31
Figura 12 - Visualização científica e visualização da informação. ....	32
Figura 13 - Representação do back-end e front-end.....	33
Figura 14 - Estrutura básica do script em PHP. ....	34
Figura 15 - Comparação entre HTML e CSS.....	36
Figura 16 - Exemplo de gráficos obtidos com a biblioteca NVD3. ....	38
Figura 17 – Representação do modelo conceitual.....	41
Figura 18 – Representação do modelo lógico. ....	42
Figura 19 - Representação da estrutura da aplicação Web.....	43

Figura 20 - Seção de visualização geral por área de conhecimento.....	44
Figura 21 – Interface do tipo <i>Checkbox</i> para escolha das áreas de conhecimento.....	45
Figura 22 - Legendas da representação visual.....	45
Figura 23 - Detalhamento dos itens selecionados. ....	46
Figura 24 - Seção comparação por estados. ....	46
Figura 25 - Comparação entre estados definida pelo usuário. ....	47
Figura 26 - Visualização por tipos de escolas. ....	48
Figura 27 - Seleção dos tipos de escolas. ....	48
Figura 28 - Visualização por região. ....	49
Figura 29 - Seleção das regiões projetadas. ....	49
Figura 30 - Seção informações gerais. ....	50
Figura 31 - Propriedade área. ....	52
Figura 32 - Legenda para as medias das notas. ....	53
Figura 33 - Legenda da cor da borda para identificar o tipo da variável. ....	53
Figura 34 - Script em PHP para importação dos dados do arquivo CSV.....	54
Figura 35 - Código da classe Conexão.....	56
Figura 36. Código da classe Dados. ....	57
Figura 37 - Código da classe Controle. ....	58
Figura 38 - Código da função "criaJsonGeralPorArea". ....	60
Figura 39 - Referencias no HTML para os <i>frameworks</i> utilizados. ....	60
Figura 40 - Código HTML utilizando o elemento <select>.....	61

Figura 41 - Navegador sem estilização do CSS. ....	61
Figura 42 - Folha de estilo para deixar os seletores e o botão mais intuitivos e visualmente.....	62
Figura 43 - Navegação com estilização do CSS.....	62
Figura 44 - Código para recuperação dos dados informados pelo usuário. ....	63
Figura 45 – Seleções por área de conhecimento. ....	63
Figura 46 - Função que converte dados em gráficos.....	64
Figura 47 - Criação de um Círculo em svg utilizando o D3.....	65
Figura 48 - Circulo criado com a D3.....	65
Figura 49 - Criando círculos com atributos.....	66
Figura 50 - Resultado da implementação dos círculos com atributos.....	66
Figura 51 - Implementação de um dos métodos de animação do d3. ....	67
Figura 52 - Criação do gráfico da sessão Visualização por tipos de escolas. ....	68
Figura 53 - Código em JSON para interpretar os dados.....	68
Figura 54 - Código alterado para ser utilizado na terceira implementação.....	69
Figura 55 - Visualização por área de conhecimento .....	71
Figura 56 - Visualização por tipo de escola .....	72
Figura 57 – Visualização por região.....	72

## **LISTA DE ABREVIATURAS E SIGLAS**

ENEM – Exame Nacional do Ensino Médio

CSS – *Cascading Style Sheets*

CRUD – *Create, Read, Update e Delete*

DDL – *Data Definition Language*

DML – *Data Manipulation Language*

HTML – *HyperText Markup Language*

IDE – *Integrated Development Environment*

MYSQL – *My Structured Query Language*

PHP – *Personal Home Page*

SGBD – Sistema de Gerenciamento de Banco de Dados

SQL – *Structured Query Language*

TIC – Tecnologias da Informação e Comunicação

UFSC – Universidade Federal de Santa Catarina

UML – *Unified Modeling Language*

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>16</b>
1.1	<i>Justificativa.....</i>	17
1.2	<i>Objetivos.....</i>	17
1.2.1	Objetivo Geral .....	17
1.2.2	Objetivos Específicos .....	17
1.3	<i>Metodologia.....</i>	18
1.4	<i>Organização do Trabalho .....</i>	18
<b>2</b>	<b>REVISÃO DA LITERATURA .....</b>	<b>20</b>
2.1	<i>Dados e Representações Visuais.....</i>	20
2.1.1	Caracterização de Dados.....	20
2.1.2	Representação Visual .....	21
2.2	<i>O que é Visualização de Dados?.....</i>	21
2.2.1	Técnicas de Análise Visual.....	22
2.2.2	Análise e Comunicação.....	30
2.2.3	Processo de Visualização .....	30
2.2.4	Visualização da Informação e Visualização Científica .....	31
<b>3</b>	<b>ARQUITETURA PARA APLICAÇÕES WEB.....</b>	<b>33</b>
3.1	<i>Arquitetura do Servidor.....</i>	33
3.1.1	A Linguagem PHP .....	34
3.1.2	O Banco de Dados MySQL .....	34
3.2	<i>A Arquitetura dos Clientes .....</i>	35
3.2.1	A Linguagem de Marcação HTML 5.....	35
3.2.2	A Linguagem de Folhas de Estilo.....	36
3.2.3	A Linguagem JavaScript.....	36
3.2.4	O Framework JQuery .....	37
3.2.5	JSON .....	37
3.2.6	A Biblioteca D3 .....	37
3.2.7	A Biblioteca NVD3 .....	38
3.2.8	O Framework Bootstrap .....	38
<b>4</b>	<b>PROJETO DA APLICAÇÃO WEB .....</b>	<b>40</b>
4.1	<i>Modelagem de dados.....</i>	40
4.1.1	Descrição do Modelo Conceitual .....	40
4.1.2	Descrição do Modelo Lógico .....	41

4.1.3	Descrição do Modelo Físico .....	42
4.2	<i>Projeto da Interface de Interação</i> .....	43
4.2.1	Visualização por área de conhecimento .....	43
4.2.2	Visualização de dados entre Estados .....	46
4.2.3	Visualização por tipos de escolas .....	47
4.2.4	Visualização por Região.....	48
4.2.5	Informações Gerais .....	50
<b>5</b>	<b>DESENVOLVIMENTO DA APLICAÇÃO WEB .....</b>	<b>51</b>
5.1	.....	51
5.2	<i>Caracterização dos Dados</i> .....	51
5.3	<i>Exploração das Técnicas de Visualização</i> .....	51
5.4	<i>Desenvolvimento das Funcionalidades do Servidor</i> .....	53
5.4.1	População dos dados.....	53
5.4.2	Desenvolvimento das Classes utilizando PHP .....	55
5.4.3	Classe conexão .....	55
5.5	<i>Desenvolvimento de Funcionalidades para o Cliente</i> .....	60
<b>6</b>	<b>RESULTADOS.....</b>	<b>70</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....</b>	<b>73</b>
<b>8</b>	<b>REFERÊNCIAS .....</b>	<b>74</b>

# 1 INTRODUÇÃO

Com o avanço tecnológico são criadas novas tecnologias capazes de solucionar problemas com alto nível de complexidade, grande volume de dados e que demandam um alto poder de processamento. Na ciência da computação isso é considerado um grande avanço tecnológico, pois traz novos recursos ou a evolução dos já existentes. Entretanto, se as quantidades de dados aumentam, de que maneira o usuário lida com isso? Qual a capacidade que um ser humano tem de assimilar uma grande quantidade de dados? É fato que o aumento em massa da informação pode gerar alguns problemas, para análise e compreensão de dados. Os dados são irrelevantes quando não se tem sentido, ou seja, para que haja informação, os mesmos devem ser interpretados e relacionados para que estejam inseridos dentro de um contexto e permitam a geração do conhecimento. Os computadores evoluem em grande escala, já os usuários, não possuem a capacidade de acompanhar tamanha evolução.

Exemplificando, pode-se imaginar um cesto com dez frutas dos mais variados tipos. É uma tarefa fácil, agrupar por tipos e compará-las pelo tamanho, massa e volume devido à pequena quantidade de frutas. Agora suponha que, ao invés de um cesto com dez frutas, tem-se um tanque com mil frutas, dos mais variados tipos. Em quanto tempo a mais um ser humano irá demorar para levantar as mesmas informações sem nenhum auxílio? Considerando esse problema, desenvolvedores cada vez mais se preocupam, e dão uma atenção especial para a visualização de dados, em que ferramentas são lançadas com o propósito de auxiliar a visualização dos dados, através de gráficos, que simplificam e esclarecem a informação para a fácil compreensão humana, que atua diretamente na tomada de decisão.

Combinando aspectos da computação gráfica, interfaces humano-computador e mineração de dados a visualização da informação permite a apresentação de dados em forma gráfica de modo que o usuário possa utilizar sua percepção visual para melhor analisar e compreender as informações. (FREITAS et al.,2001)

Este trabalho visa demonstrar a importância da visualização da informação, essencialmente sobre um grande volume de dados. Para isso, a base de dados escolhida no desenvolvimento da ferramenta de visualização foi o micro dados do ENEM, que conta com mais de 7 milhões de linhas.



## **1.1 Justificativa**

Com o crescimento de dados na era da informação e comunicação, torna-se mais difícil manipular e fazer o tratamento dos mesmos. A proposta deste trabalho é desenvolver uma ferramenta para facilitar o processo de visualização e análise dos dados do ENEM, de uma maneira clara e objetiva, que facilita o processo de assimilação do ser humano. Assim, transformando toda informação em conhecimento com a finalidade de melhorar o processo da tomada de decisão.

Atualmente, não há muitas ferramentas interativas que permitam a visualização e análise destes dados na área educacional, por esse motivo, surgiu a ideia de desenvolver uma aplicação Web capaz de representar graficamente os dados do ENEM. Além de despertar curiosidades, a visualização dessas informações, poderá também servir para as instituições de ensino verificarem o desempenho de seus alunos nos seus respectivos estados, e assim, analisar qual área dominam e qual deixam a desejar e, talvez ainda, tomar medidas para melhorar o desempenho em exames futuros.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

O presente trabalho tem como objetivo principal desenvolver uma ferramenta Web para visualização de dados do ENEM para facilitar o processo de análise dos dados e tomada de decisões.

### **1.2.2 Objetivos Específicos**

Para que seja possível alcançar o objetivo principal, são requeridos os seguintes objetivos específicos:

- Estudar os diferentes algoritmos de visualização de dados.
- Analisar as principais tecnologias utilizadas em visualização de dados.

- Classificar as tecnologias utilizadas para o desenvolvimento da aplicação.
- Desenvolver o projeto da aplicação Web, a interface de interação e modelagem de dados para visualização.

### **1.3 Metodologia**

Para a elaboração deste trabalho, foi realizada uma pesquisa aplicada e tecnológica. O trabalho contou com um estudo geral na área de visualização de dados. Para implementação da aplicação Web foram escolhidas tecnologias que mais se adequam e otimizam o processo de visualização dos dados do ENEM.

Quanto aos dados do ENEM, foram obtidos do portal de dados do Governo Federal, onde são disponibilizados dados abertos, das mais diversas áreas públicas.

No desenvolvimento da aplicação Web foram utilizados softwares livres. Optou-se pela linguagem de programação PHP e adotou-se o ambiente de desenvolvimento Web conhecido como WAMP. Para gerenciamento de dados, foi utilizado o banco de dados MySQL, que apresentou maior afinidade em relação a outros bancos de dados, pois este possui um excelente desempenho e uma grande base de documentação. Elaborou-se também, a modelagem de dados conceitual, lógica e relacional para a criação do banco de dados e suas tabelas, relacionamentos e colunas. E, para visualização de dados, adotou-se a biblioteca D3.

### **1.4 Organização do Trabalho**

O segundo capítulo trata-se de uma revisão da literatura, onde o objetivo é contextualizar e apoiar o tema deste trabalho que aborda a visualização de dados.

O terceiro capítulo descreve as principais tecnologias utilizadas ao decorrer deste trabalho que contribuíram para o desenvolvimento da aplicação Web.

O quarto capítulo apresenta de uma forma geral o projeto com todas as suas partes, como a modelagem de dados, interface de interação e suas respectivas funcionalidades.

O quinto capítulo descreve de forma detalhada o desenvolvimento da aplicação, incluindo os principais algoritmos desenvolvidos e modelos utilizados para a implementação da ferramenta Web.

O sexto capítulo apresenta os resultados alcançados com este trabalho dentro do tema proposto no desenvolvimento da aplicação Web para visualização de dados do ENEM.

Por fim, o sétimo capítulo apresenta as considerações finais, ressaltando a importância de ferramentas para a visualização da informação na atualidade e as desvantagens que a sua ausência pode trazer ao processo de análise e tomada de decisão.

## **2 REVISÃO DA LITERATURA**

A revisão da literatura tem uma função muito importante em trabalhos acadêmicos. Ela é quem inicia o processo de pesquisa bibliográfica e tem a finalidade de abordar de forma consistente o assunto que o trabalho apresenta.

A revisão da literatura idealiza o estado atual do conhecimento sobre o dado tema e contribui para o desenvolvimento do conhecimento. Conforme Cardos (2010) “cada investigador analisa minuciosamente os trabalhos dos investigadores que o precederam e, só então, compreendido o testemunho que lhe foi confiado, parte equipado para a sua própria aventura” (p. 7).

### **2.1 Dados e Representações Visuais**

Os dados podem ser considerados pela observação documentada ou resultado de medições, que descrevem um determinado fenômeno ou entidade que são objetos de estudos e análises (FREITAS, 1995).

#### **2.1.1 Caracterização de Dados**

Existem diferentes critérios para caracterizar os dados levando em consideração que os mesmos se correspondem a atributos. A identificação destas, permitem a escolha de uma técnica de visualização que melhor se adeque ao conjunto de dados.

Para Freitas (1995), o primeiro critério a ser identificado para caracterizar os atributos é a sua classe de informação que representa. Ou seja, deve-se verificar a qual classe, cada atributo melhor se enquadra. Outro critério a se observar é o seu tipo, deve-se identificar quais valores o atributo poderá assumir, por exemplo: alfanumérico, inteiros reais, símbolos, dentre outros. Os dados também podem se caracterizar pela sua dimensão e a natureza de domínio onde estão definidos, podendo ser unidimensionais, bidimensionais, tridimensionais e até n-dimensionais.

### **2.1.2 Representação Visual**

Também conhecida como gráfico, a representação visual corresponde às figuras, ou imagens que são utilizadas para representar um subconjunto de dados, que estão sob análise. Entre inúmeros modelos de gráficos utilizados para apresentar os dados, destacam-se o de barras, frequência e pontas, que permitem analisar as relações entre atributos.

A representação visual não se limita em apenas modelos tradicionais, ela se torna flexível para elaboração gráfica de novas maneiras de expressar os atributos. Entretanto, existem alguns pontos relevantes que devem ser observados: transmitir de forma clara e direta e facilitar a compreensão sem exigir grandes esforços. É importante aqui, ressaltar que, a representação visual não se preocupa com a estética perfeita, mas sim, em tornar a interface agradável, que transmita a informação de maneira rápida e direta, facilitando a análise e compreensão.

## **2.2 O que é Visualização de Dados?**

Visualização é um processo cognitivo, que tem como função: gerar uma imagem mental de alguma coisa; internalizar uma compreensão. Seu objetivo principal, não é apenas transmitir gráficos, mas passar a compreensão necessária para novos descobrimentos, explicações e tomadas de decisão.

Visualização da Informação é ideal para análise exploratória de dados. Nossos olhos são naturalmente atraídos por tendências, padrões e exceções que seriam difíceis ou impossíveis de se achar utilizando abordagens tradicionais, tais como tabelas ou textos. Durante a exploração dos dados, até mesmo os melhores estatísticos, muitas vezes colocam os seus cálculos de lado por um momento e deixam os seus olhos conduzirem a busca. (FEW, 2009)

### 2.2.1 Técnicas de Análise Visual

Em geral os dados não possuem uma representação direta, óbvia e natural, o que contribui para a utilização de uma boa técnica de análise visual, essa por sua vez utiliza representações visuais para representar graficamente os dados (LUZZARDI, 2003).

Card, Mackinlay e Shneiderman, (1999) citam que “o uso de representações visuais de dados interativas com o suporte do computador permite ampliar o conhecimento”. Entretanto, deve-se escolher uma técnica que mais se adeque às características dos dados que serão analisados, pois o objetivo da visualização é facilitar a compreensão e só será possível com a utilização de uma boa técnica de Análise Visual.

Dentre tantas técnicas existentes, geralmente há uma busca de inspiração em objetos do mundo real por parte dos autores, para estruturar o conjunto de dados. Outros optam por formas geométricas, por se constituírem de forma simples e de fácil representação. Estas podem ser representadas em uma, duas ou três dimensões (1D, 2D ou 3D), não necessariamente de acordo com a dimensão do espaço de informação.

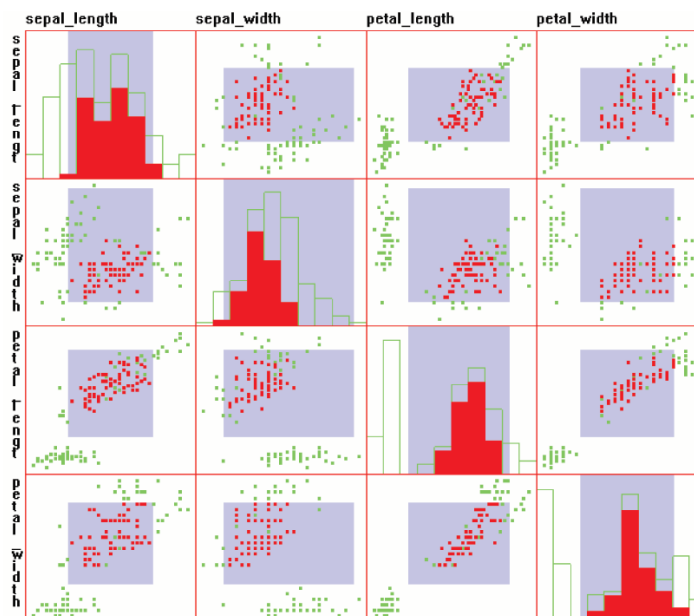
A seguir, algumas técnicas são apresentadas, demonstrando uma representação eficaz sobre um determinado conjunto de dados. Elas podem se basear em gráficos, informações quantitativas, hierárquicas, exploração de números dentre outros.

#### 2.2.1.1 Dados Multivariados

Para Ward, Grinstein e Kein (2015), os dados multivariados são aqueles que geralmente não possuem um atributo espacial explícito. Sobre este conjunto de dados, pode-se utilizar as seguintes técnicas:

- **Técnicas Baseadas em Pontos:** Esta técnica utiliza registro em um espaço de dados ‘n’ dimensional para ser projetado em um espaço ‘k’ dimensional arbitrário. Baseia-se na projeção de pontos de uma dimensão elevada para o espaço 2D ou 3D, como se pode observar a ilustração a seguir (Figura 1).

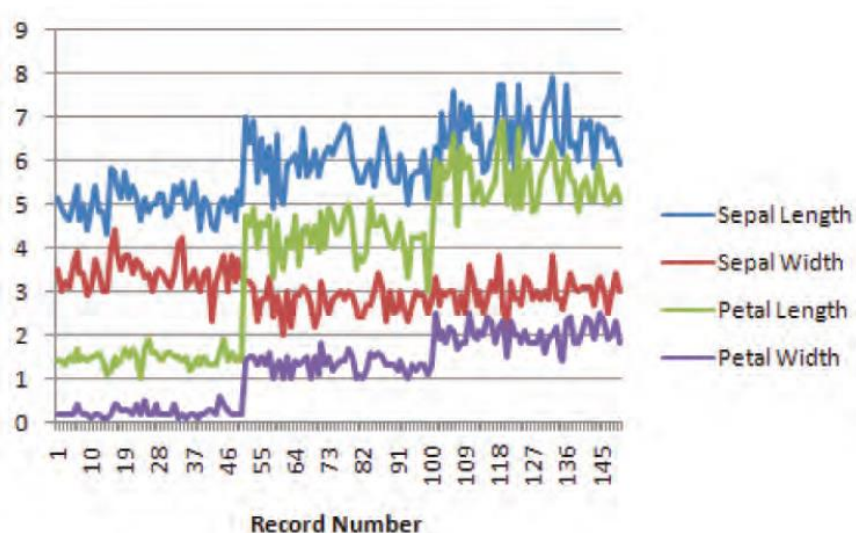
Figura 1 - Técnica de visualização de dados baseada em pontos.



Fonte: Ward, Grinstein e Kein (2015).

- **Técnicas Baseadas em Linhas:** Esta técnica, representada na imagem abaixo, utiliza pontos de um registro particular ou no espaço 2D, onde estes estão conectados por linhas retas ou por linhas curvas, como exemplo os gráficos de linhas e coordenadas paralelas (Figura 2).

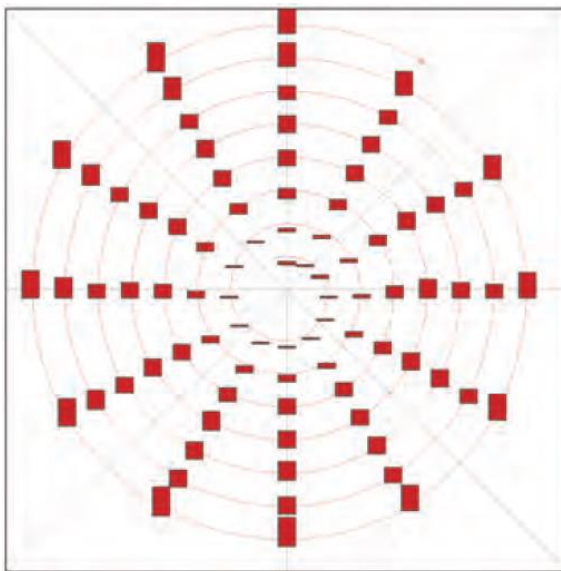
Figura 2 - Técnica de visualização de dados baseada em linhas.



Fonte: Ward, Grinstein e Kein (2015).

- **Técnica Baseada em Regiões:** Esta por sua vez, utiliza preenchimentos por polígonos para transmitir informação de valores (Figura 3), levando em consideração sua forma, cor dentre outros atributos. Um exemplo muito conhecido desta técnica é o gráfico de polígonos.

**Figura 3 - Técnica de visualização de dados baseada em regiões.**



**Fonte: Ward e Lipchak (2000).**

- **Combinações de Técnicas:** Como seu nome propriamente diz, esta técnica nada mais é do que a combinação das características de duas ou mais técnicas apresentadas anteriormente. Por exemplo, a criação de glifos (figuras que dão um tipo de característica particular a um símbolo específico) é considerada uma combinação de técnicas. A Figura 4 representa esta técnica.



Figura 4 - Combinações de técnicas sobre dados multivariados.



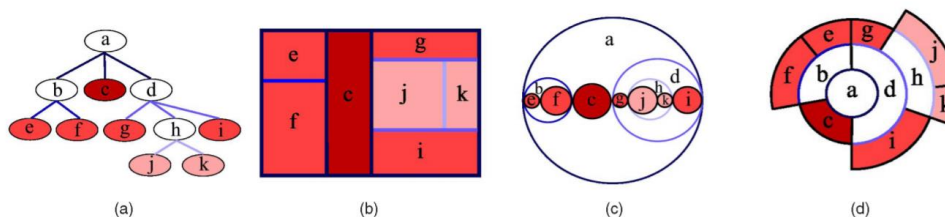
Fonte: Ward (2002).

### 2.2.1.2 Hierarquias e Árvores

Dados com característica hierárquicas, são considerados um repositório de dados onde existe relação com subitens de dados. Esta relação pode ser vista como uma imposição de ordem e podem ser parentes ou descendentes de outros itens do repositório. A hierarquia geralmente é representada como árvore. Os tópicos a seguir descrevem os dois principais métodos de representação desta característica de dados. (WARD; GRINSTEIN; KEIN, 2015)

- **Métodos de Preenchimento de Espaço.** Este método utiliza o espaço para codificar outras variáveis de itens de dados. É realizada a junção de objetos ao invés dos mesmos serem ligados por arestas. Esta técnica é conhecida como *TreeMaps*, e geralmente utiliza representações retangulares e radiais para a abordagem. A Figura 5 descreve este método.

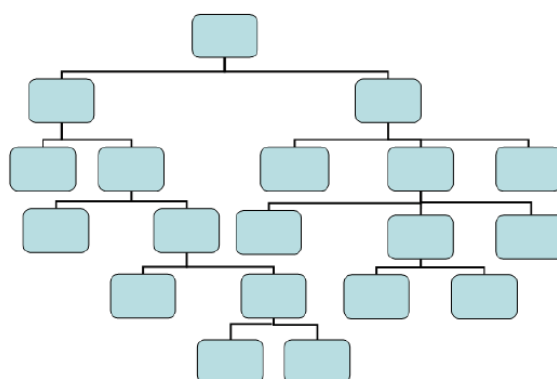
**Figura 5 - Método de preenchimento de espaço.**



Fonte: Caserta e Sendra (2011).

- Métodos de Não-Preenchimento de Espaço:** Ao contrário do método anterior, este é o mais comum utilizado para visualização de relacionamentos hierárquicos, e também é conhecido como o diagrama de ligação de nós. É utilizado, normalmente, apenas para representação de grau parentesco, onde pode possuir diversos níveis hierárquicos, como representado na Figura 6.

**Figura 6 - Método de não preenchimento de espaço.**



Fonte: Ward, Grinstein e Kein (2015).

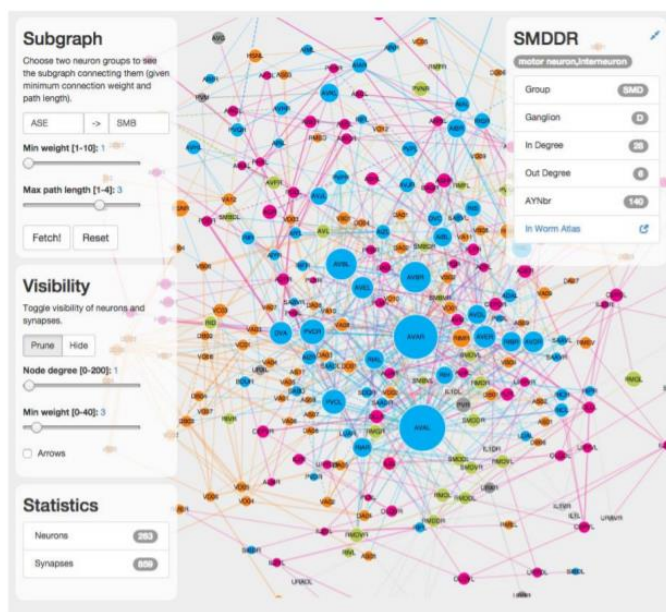
### 2.2.1.3 Grafos e Redes

Estes tipos de dados são definidos por suas características de relação baseadas em grafos, que são dois ou mais vértices que estão unidos por uma aresta, podem ser direcionais ou não. O grau de um vértice é o número de vértices conectados a ele. Arestas de grafos

podem ter valores (pesos). A seguir são representadas duas técnicas mais comuns utilizadas para dados com esta característica. (WARD; GRINSTEIN; KEIN, 2015)

- **Restrições Baseadas em Forças:** Em geral esta técnica utiliza um sistema com base em molas. Onde suas arestas possuem as propriedades de atração da gravidade. E seus vértices possuem partículas carregadas, ou seja, repulsão. Normalmente, adota-se uma equação para calcular as forças, e de modo interativo ao sofrer uma ação sobre, é recalculada a posição dos vértices (Figura 7).

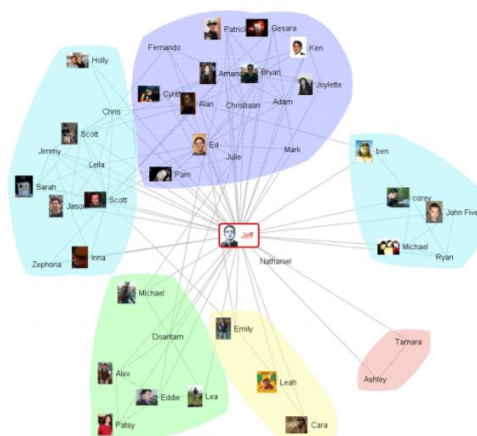
**Figura 7 - Técnica de redes e grafos com restrições baseadas em forças.**



**Fonte: Worm (2014).**

- **Baseados em atributos:** Esta técnica pode utilizar subgrupos para classificar as conexões de acordo com suas características, como se pode observar na Figura 8.

**Figura 8 - Técnica de grafos e redes baseada em atributos.**



**Fonte: Herr e Boyd (2005)**

#### **2.2.1.4 Textos e Documentos**

Este conjunto de dados pode ser considerado por ser do tipo literal, como uma cadeia de caracteres, e definido pelo conjunto dos objetos: palavras, sentenças, parágrafos e documentos. Assim como os demais tipos de dados, também existem métodos utilizados para sua representação visual. Algumas características a serem analisadas podem considerar a associação com meta-dados como por exemplo, autor, data, tamanho, citações, etc. A seguir são apresentadas algumas técnicas utilizadas para dados que possuem estas características. (WARD; GRINSTEIN; KEIN, 2015)

- **Visualização da Recuperação da Informação:** Esta técnica permite recuperar a informação de textos e documentos, como por exemplo, através de consultas é realizada a representação visual, de acordo com o atributo do documento a ser analisado. Estes atributos podem incluir a frequência dos termos no documento, a distribuição, a sobreposição, o comprimento do documento e etc. A Figura 9 demonstra a utilização desta técnica, onde a frequência apresentada por cada palavra é proporcionalmente relacionada ao seu tamanho no espaço da visualização.



### **2.2.2 Análise e Comunicação**

Ao comparar análise e comunicação, percebe-se que ambas se complementam. Primeiramente, realiza-se a comunicação, objetivando o entendimento do que foi comunicado para tomar uma decisão. Basicamente, elas possuem muito em comum, não se pode analisar algo, que não foi comunicado, devido a inexistência da informação.

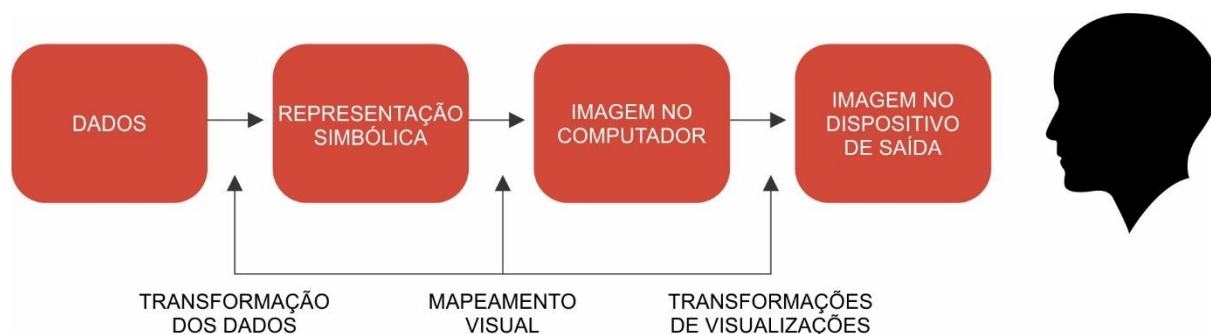
Após fornecer os dados, é realizado o entendimento, comparação, decisão, julgamento e avaliação dos dados. E o que contribui para tornar esta afirmação verdadeira, responsabilidade da comunicação, é informar aos outros de uma maneira mais efetiva para que os mesmos compreendam e ajam sobre o entendimento dado. Geralmente, a análise que visa uma resolução de problemas, é muito utilizada em estatísticas, banco de dados, mineração de dados e aprendizagem de máquina.

A ideia de uma comunicação efetiva ocorre pela possibilidade de tornar a visualização útil na análise exploratória de dados. Ou seja, quando não se sabe o que realmente se está procurando, ou não se possui as questões de antemão e se deseja saber o que questionar. Pode-se concluir que, a visualização é utilizada para comunicar ideias, influenciar, explicar e persuadir. Ela conta com auxílios fundamentais, de extrema importância para sua eficácia, que nada mais são do que os elementos visuais, que podem servir como uma evidência ou suporte, e frequentemente são utilizados para substituir palavras, resumir, agregar, unir e explicar, dando maior objetividade à comunicação.

### **2.2.3 Processo de Visualização**

O processo de visualização é definido pelo mapeamento de elementos através de gráficos, que são exibidos em um dispositivo de saída. As principais etapas da visualização que são consideradas no uso de uma determinada técnica podem ser identificadas por meio de modelos de referência de Visualização, como o modelo apresentado por Card (1999) na Figura 11.

**Figura 11 – Etapas do processo de visualização baseada no modelo de referência de Card (1999).**



**Fonte: Autores.**

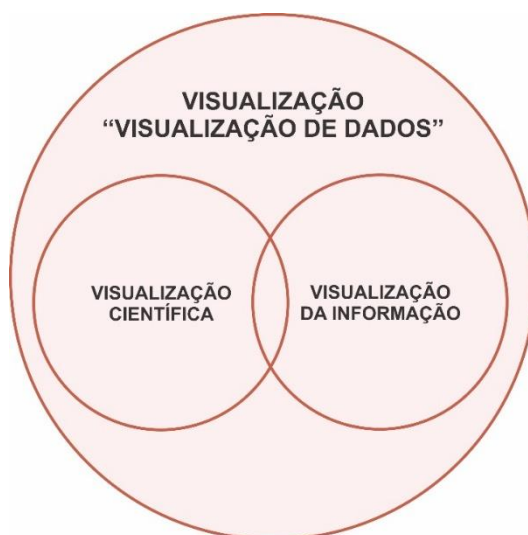
O modelo apresentado por Card (1999), descreve o processo de visualização como uma sequência de mapeamentos de dados tabelados para uma representação visual. Possibilitando a interação do usuário com o espaço de informação, permitindo assim, obter o que é conhecido por cristalização do conhecimento.

Logo, entende-se o processo de visualização, como um processo de mapeamento de dados e informações através de representação visual e mecanismos de interação. A computação auxilia diretamente na ampliação do conhecimento, oferecendo recursos que visam melhorar a visualização. (CARD, 1999)

#### **2.2.4 Visualização da Informação e Visualização Científica**

A visualização de dados é um conjunto formado pela Visualização Científica e a Visualização da Informação, como mostra a Figura 12.

**Figura 12 - Visualização científica e visualização da informação.**



**Fonte: Autores.**

A definição de visualização científica é dada pela representação visual de dados do tipo científicos, onde são usualmente físicos em natureza ao invés de abstratos. A representação de órgãos presentes no corpo humano modelados a partir de imagens médicas, é um exemplo de visualização científica.

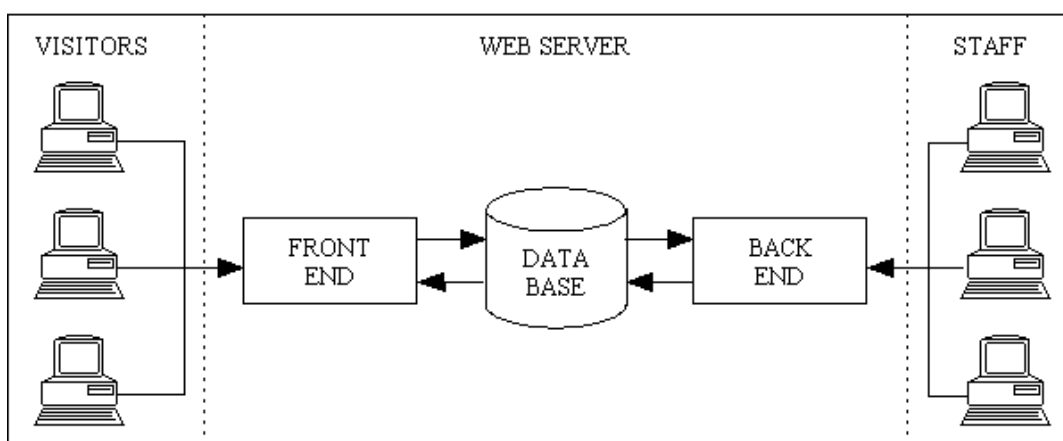
Já, a Visualização da Informação é a representação de dados abstratos, com a finalidade de aumentar o conhecimento. Ou seja, auxilia diretamente no processo de construção do conhecimento e é muito útil para a tomada de decisão. A representação de estatísticas de números ligados às informações na Internet é um exemplo de Visualização da Informação.



### 3 ARQUITETURA PARA APLICAÇÕES WEB

Neste capítulo serão apresentadas as principais tecnologias que serviram de utilidade para o desenvolvimento da aplicação de visualização Web. Basicamente, as tecnologias são divididas em duas: a parte do *front-end*, que realiza o papel de interface da interação do usuário, e o *back-end*, que possui o controle do banco de dados e da própria aplicação. Lamim (2014) apresenta o funcionamento e as interações entre o *front-end* e o *back-end*, como se pode ver na Figura 13.

Figura 13 - Representação do back-end e front-end.



Fonte: Lamin (2014).

#### 3.1 Arquitetura do Servidor

O servidor é um computador de alto poder de processamento e armazenamento, utilizado para fornecer serviços aos clientes que o acessarão.

A arquitetura do servidor também é conhecida como *back-end*. O *back-end* é um termo utilizado para representar a etapa final de um processo, na Web, ele é conhecido por compor as funcionalidades internas do servidor. Basicamente é composto por dois elementos: aplicação e banco de dados. A aplicação, é o código programado para ser executado. Já, o banco de dados tem a função de armazenar dados, que são utilizados na aplicação.

### 3.1.1 A Linguagem PHP

Uma das linguagens de script utilizada para desenvolvimento da aplicação Web foi o PHP (acrônimo recursivo do inglês: *Hipertext Processor*), que se trata de uma linguagem de script *Open Source*. Geralmente é utilizada para desenvolvimento Web, devido a possibilidade da mesma se embutir ao HTML delimitado pelas *tags* de início e fim “<php e ?>”, como no exemplo apresentado na Figura 14.

**Figura 14 - Estrutura básica do script em PHP.**

```
1 <html>
2   <head>
3     <title>PHP Exemplo</title>
4   </head>
5   <body>
6     <?php echo "<p>Psiu! Sou um script.</p>"; ?>
7   </body>
8 </html>
```

**Fonte: Autores.**

O PHP é compatível com a maioria dos sistemas operacionais e também é suportado por grande parte dos servidores atuais, o que dá a liberdade para a escolha desejada. Também, pode-se optar pela programação estruturada, orientada à objetos ou mista.

Apesar da linguagem ser embutida no HTML, não se limita apenas à geração da mesma. Com o PHP, pode-se gerar imagens, arquivos em PDF e até mesmo, animações em Flash dinâmicos.

O PHP é amplamente suportado por vários banco de dados, onde é necessário utilizar extensões específicas para que a página web consulte o banco de dados com facilidade.

### 3.1.2 O Banco de Dados MySQL

O banco de dados MySQL, foi o escolhido para este estudo por ser utilizado por uma grande comunidade e por ser distribuído por uma licença de software livre.

O MySQL é um SGBD (Sistema Gerenciador de Banco de Dados) que utiliza a linguagem para sua interface. Desenvolvida por David Axmark, Allan Larsson e o finlandês Michael Widenius, na Suécia, este banco de dados é popularmente utilizado e considerado por muitos, um banco de fácil acesso, com uma interface simples e com capacidade de ser compatível com os mais diversos sistemas operacionais.

Outro motivo por muitos optarem o MySQL, dentre a tantos concorrentes, é porque ele é protegido por uma licença de software livre, desenvolvida pela GNU.

## **3.2 A Arquitetura dos Clientes**

O cliente pode ser considerado como o usuário que irá acessar os recursos ou serviços disponíveis no servidor. Para que o cliente possa fazer esses requerimentos, existem alguns componentes que auxiliam neste processo. A arquitetura do cliente também é conhecida como *front-end*. A seguir, serão apresentados os componentes principais, que compõem a arquitetura do cliente.

### **3.2.1 A Linguagem de Marcação HTML 5**

O HTML (*Hypertext Markup Language*) trata-se de uma linguagem de marcação criada em 1991, para ser um padrão de comunicação e troca de documentos na Web. A linguagem faz uso de conteúdos multimídias (texto, imagens, vídeos e áudios) para criação de páginas Web e comunicar determinada informação (W3C ESCRITÓRIO BRASIL, 2016).

Baseada na utilização de hipertexto, esta linguagem de marcação possibilita a agregação de blocos de informação, como os conteúdos multimídia, que serão acessados através de hyperlinks ou links. Este conceito permite que o acesso possa atingir a alta quantidade de informações, que estão peças e conectadas, que juntos formam uma grande rede de informação, a internet (W3C ESCRITÓRIO BRASIL, 2016).

A estrutura que compõem as páginas em HTML ocorre pela utilização de *tags*, que servem para marcar os objetos inseridos e detalhar como eles serão exibidos. As *tags* são termos colocados entre colchetes angulares, que delimitam o início e fechamento dos objetos.

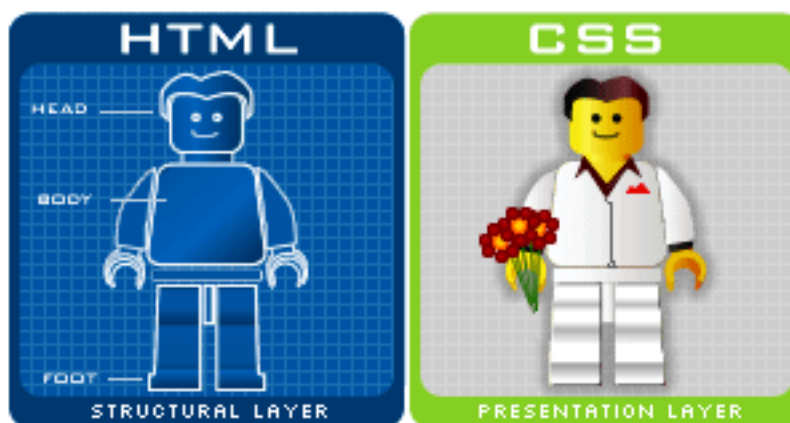
Para leitura de uma página em HTML, necessita-se de um navegador web (*web browser*), geralmente conhecido como cliente, que tem a função de interpretar e representar, visualmente, o conteúdo presente no código.

### 3.2.2 A Linguagem de Folhas de Estilo

A linguagem CSS (do inglês *Cascading Style Sheets*) é uma folha de estilo em cascata que serve para estilizar a maneira como as páginas em HTML serão exibidas na tela. Ou seja, ela pode garantir uma formatação homogênea e uniforme em páginas de um site, com um visual elegante. Sua criação surgiu com o intuito de separar a estrutura do documento de sua aparência (GRANNEL, 2007).

A Figura 15 apresenta a relação da funcionalidade do HTML e CSS, em que toda a estrutura é responsabilidade do HTML e o design do CSS.

**Figura 15 - Comparação entre HTML e CSS.**



Fonte: Navarro (2016).

### 3.2.3 A Linguagem JavaScript

JavaScript (JS) é uma linguagem de programação interpretada que disponibiliza vários recursos de interface gráfica, como botões, campos e seletores, necessários para desenvolvimento de páginas de internet, com mais interatividade. A mesma se encontra disponível nos navegadores de internet, que fazem parte da arquitetura do cliente.

Com a sintaxe semelhante a apresentada pela linguagem C, a JS permite ainda, a integração e modificação de forma dinâmica de conteúdos e visuais, dos elementos que a compõe (BORTOLOSSI, 2012).

### 3.2.4 O Framework JQuery

O jQuery é uma biblioteca em JavaScript desenvolvida por John Resig, com o objetivo de simplificar a criação de efeitos visuais e de interatividade em sites da Web. Esta foi a diretriz que levou o autor a sua criação, a simplicidade no desenvolvimento de aplicações.

Proporcionando uma criação de scripts de forma intuitiva, o jQuery permite minimizar a quantidade de linhas no código mantendo o mesmo efeito. Por exemplo, dez linhas de código com o jQuery, pode facilmente apresentar os mesmos efeitos de um script de 40 linhas de código desenvolvido através de JavaScript.

### 3.2.5 JSON

JSON (*JavaScript Object Notation*) é um modelo para armazenamento e transmissão de dados baseado em um subconjunto da linguagem de programação JavaScript. JSON consiste em um formato de texto leve e completamente independente de linguagem, pois usa convenções que são familiares às linguagens C e familiares, incluindo PHP, C++, C#, Java, JavaScript, Perl, Python e entre outras (JSON, 2016).

Com este formato é possível armazenar e interpretar os dados de modo simples e rápido. Desta forma, pode se dizer que JSON é o formato ideal para a comunicação entre o *back-end* e *front-end* quando se trata de transferência de um número considerável de dados como no caso da aplicação em questão.

### 3.2.6 A Biblioteca D3

O D3 (do inglês *Data Driven Documents* – D3) é uma biblioteca da linguagem JavaScript utilizada para a visualização de dados baseados nos padrões da Web. Com a utilização de SVG, Canvas e HTML, busca facilitar a compreensão dos dados através da combinação da interação com as técnicas de visualização. A mesma foi projetada para ser

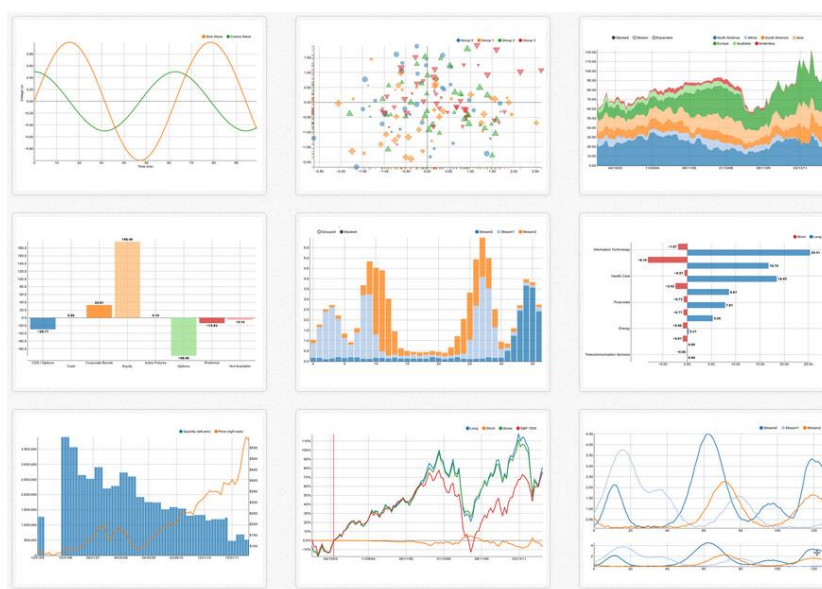
compatível com diversos navegadores e dar liberdade para projetar a interface visual ideal para seus dados.

O foco da D3 não é proporcionar uma estrutura com todos os recursos imagináveis, mas sim, uma ferramenta eficiente que resolva o problema de manipulação de documentos presentes em uma base de dados. Oferecendo uma extraordinária flexibilidade, o D3 proporciona com a mínima sobrecarga, recursos para o rápido desenvolvimento, apoiando grandes conjuntos de dados e comportamentos dinâmicos de interação e animação.

### 3.2.7 A Biblioteca NVD3

A NVD3 é uma biblioteca baseada na D3 que disponibiliza gráficos prontos reutilizáveis e componentes de gráfico para D3. Os gráficos são de fácil implementação e totalmente personalizáveis. Na Figura 16 é possível visualizar algumas implementações utilizando o NVD3.

**Figura 16 - Exemplo de gráficos obtidos com a biblioteca NVD3.**



Fonte: NVD3 (2016).

### 3.2.8 O Framework Bootstrap

O framework Bootstrap é popularmente conhecido por ser projetado para auxiliar no desenvolvimento *front-end*, fazendo com que este processo seja mais rápido e fácil. O

framework presta suporte a HTML, CSS e JS. Possibilita a criação de projetos responsivos, para serem utilizados nos dispositivos móveis.

O que leva a eficiência deste, vem do fato de apenas um *framework* suportar qualquer dispositivo. De uma maneira fácil e eficiente, o Bootstrap adapta seus projetos Web com um simples código, de smartphones para tablets e para desktops através do recurso conhecido como *media queries* do CSS. Além disso, apresenta uma excelente documentação para elementos HTML, componentes CSS e HTML personalizados e *plug-ins* jQuery.

Segundo informações presentes em seu site oficial, “o Bootstrap usa CSS tradicional, mas seu código fonte utiliza os dois pré-processadores CSS mais populares, Less e Sass” (BOOTSTRAP, 2016).

## **4 PROJETO DA APLICAÇÃO WEB**

O projeto da aplicação iniciou-se através da necessidade de visualizar dados relevantes do ENEM como, número de inscrições, médias de notas e idade e comparações de rendimento de estados por tipo de escola ou área de conhecimento de uma forma eficiente e informatizada, levando em conta as técnicas de visualização de dados apresentadas nos tópicos anteriores. Diante disto, as etapas com a modelagem de dados e suas especificações são especificados nas próximas seções.

### **4.1 Modelagem de dados**

Os dados referentes ao ENEM foram obtidos diretamente no Portal Brasileiro de dados abertos ([dados.gov.br](http://dados.gov.br)). O portal é uma ferramenta disponibilizada pelo governo brasileiro com a função de apresentar para todos os dados e informações públicas. O mesmo preza pela organização e simplicidade, de forma em que se possa encontrar facilmente os dados e informações necessárias para a sua utilização.

Com objetivo de documentar e normalizar a representação do ambiente da aplicação, foi necessário a elaboração de alguns modelos de dados e interface de interação os quais serão descritos a seguir.

#### **4.1.1 Descrição do Modelo Conceitual**

O protótipo conceitual para a estruturação de dados se baseia na abstração da tecnologia, que visa solucionar a proposta de negócio de modo mais subjetivo possível. Com isso, foi utilizado uma modelagem com base em diagramas de entidade-relacionamento. O modelo apresentado na Figura 17 descreve o conceito da aplicação, que possui somente uma tabela para representar o candidato, com seus atributos, que incluem número de inscrição, notas obtidas, tipo de escola frequentada e estado que reside.



**Figura 17 – Representação do modelo conceitual.**



**Fonte: Autores.**

#### **4.1.2 Descrição do Modelo Lógico**

A aplicação Web, adotou o modelo lógico apresentado na Figura 17. O mesmo tem a finalidade de representar, por exemplo, os relacionamentos entre as tabelas, as chaves primárias e os domínios de cada campo. No caso do VisdadosEnem, como os dados estão unificados em apenas uma tabela, o modelo não apresenta qualquer tipo de relacionamento com outras tabelas. Pode-se observar na Figura 17, a identificação da chave primária, e os tipos de cada atributo com seu tamanho.

**Figura 18 – Representação do modelo lógico.**

CANDIDATO	
	NU_INSCRICAO: BIGINT(10)
	UF_RESIDENCIA: VARCHAR(2)
	TIPO_ESCOLA: VARCHAR(1)
	NOTA_CN: BIGINT(9)
	NOTA_LC: BIGINT(9)
	NOTA_CH: BIGINT(9)
	NOTA_MT: BIGINT(9)
	NU_NOTA_REDACAO: BIGINT(9)
	IDADE: INTEGER

**Fonte: Autores.**

#### **4.1.3 Descrição do Modelo Físico**

Com a geração do banco de dados, logo, é criada a tabela com seus atributos, seguindo a definição do modelo lógico. A seguir, é descrito o script de criação da estrutura de armazenamento permanente do banco de dados:

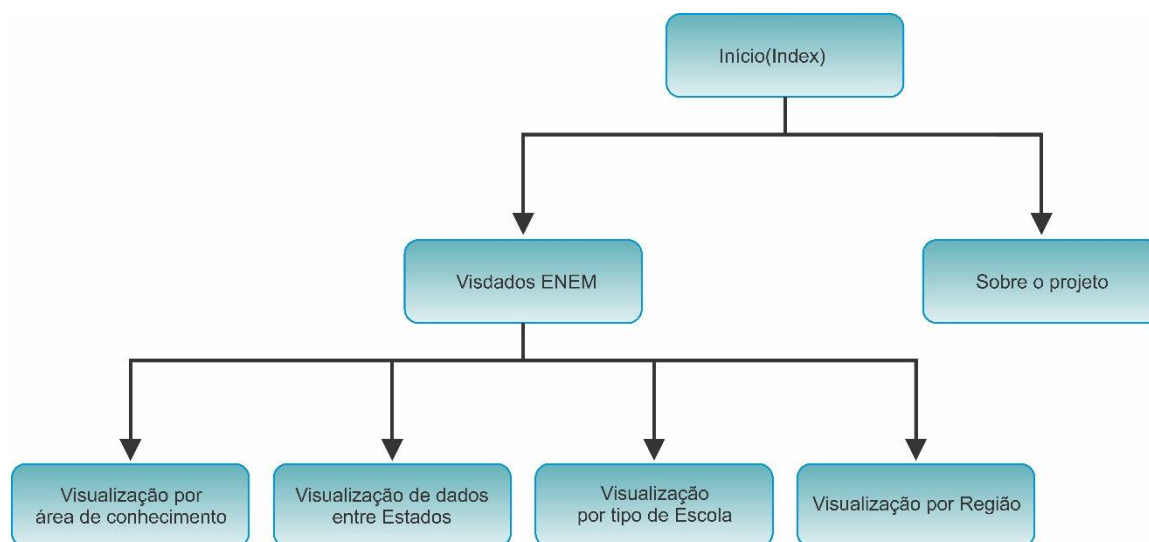
```
CREATE TABLE `TCC`.`ENEM_2014` (
  NU_INSCRICAO BIGINT(10) PRIMARY KEY,
  UF_RESIDENCIA VARCHAR(2),
  NOTA_CN BIGINT(9),
  NOTA_CH BIGINT(9),
  NOTA_LC BIGINT(9),
  NOTA_MT BIGINT(9),
  NU_NOTA_REDACAO BIGINT(9),
  TIPO_ESCOLA VARCHAR(1)
)
```

## 4.2 Projeto da Interface de Interação

O projeto da interface apresenta uma estrutura onde o usuário tem acesso facilitado a todas as funcionalidades da aplicação como é descrito na Figura 19. A interface possui as seguintes seções: Visualização por área de conhecimento, Visualização de dados entre estados, Visualização por tipo de escolas, Visualização por região, Contato e Sobre.

Para proporcionar um ambiente com mais interação do usuário, buscou-se implementar filtros para que o mesmo possa definir as comparações realizadas em cada funcionalidade. Ou seja, o usuário irá selecionar o que quer visualizar, como por exemplo o tipo de escola, o estado e etc.

**Figura 19 - Representação da estrutura da aplicação Web.**



**Fonte: Autores.**

### 4.2.1 Visualização por área de conhecimento

Conforme apresentado na Figura 20, pode-se visualizar as notas obtidas em todos os estados nos diferentes tipos de escola. Possibilitando assim, comparar as notas e analisar os estados que mais se destacam e os que menos se destacam no cenário nacional. Também é

possível observar o rendimento de cada tipo de escola em relação a nota média do Brasil, e comparar, de modo geral, a quantidade de inscritos e suas notas obtidas.

**Figura 20 - Seção de visualização geral por área de conhecimento.**



**Fonte: Autores.**

A interface oferece uma interação com o usuário, permitindo que o mesmo escolha as médias de notas por áreas através dos campos *checkbox* (Figura 21). Esta opção permite um leque maior de combinações possíveis, podendo assim, explorar detalhadamente a análise sobre o conjunto de dados.

**Figura 21 – Interface do tipo *Checkbox* para escolha das áreas de conhecimento.**

Selecione a(s) área(s) de conhecimento:

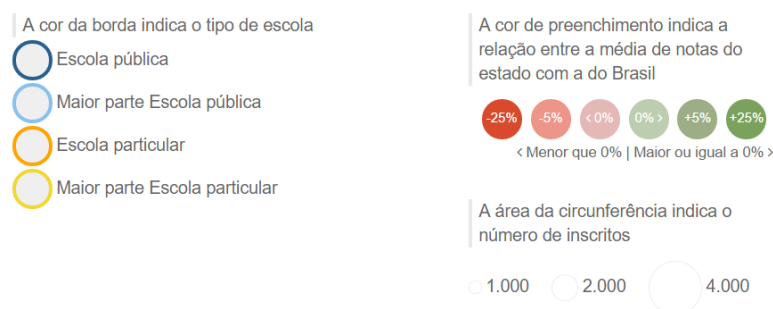
- ☒ Ciências da Natureza
- ☒ Ciências Humanas
- ☒ Linguagens e Códigos
- ☒ Matemática
- ☒ Redação

**Fonte Autores.**

Para analisar as médias da redação por estado, logo, deve-se marcar o campo de Redação e desmarcar os demais. A partir disso, os resultados serão apresentados apenas com as médias de redação.

O canto superior direito da interface, apresentado na Figura 22, possui legendas que associam os atributos do objeto com o tipo de informação representada. Esta parte da interface é de extrema importância, pois sem ela seria impossível compreender a comunicação da visualização. Por este motivo, foi cuidadosamente detalhada para esclarecer ao máximo as informações da projeção.

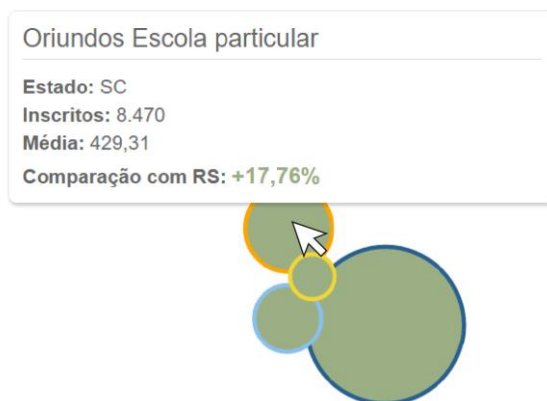
**Figura 22 - Legendas da representação visual.**



**Fonte: Autores.**

Quando o cursor do mouse está sobre uma das bolhas (tipo de escola), é projetada uma nova janela sobre a tela principal, como mostra a Figura 23. Esta janela tem o papel de descrever de forma quantitativa os números obtidos pela escola selecionada.

**Figura 23 - Detalhamento dos itens selecionados.**



Fonte Autores.

#### 4.2.2 Visualização de dados entre Estados

Esta funcionalidade, representada na Figura 24, tem como objetivo apresentar o aproveitamento de um estado em comparação a outro estado ou ao Brasil. Por exemplo, se comparar o estado A com B, a representação resultará no aproveitamento do estado A com base nas médias obtidas por B.

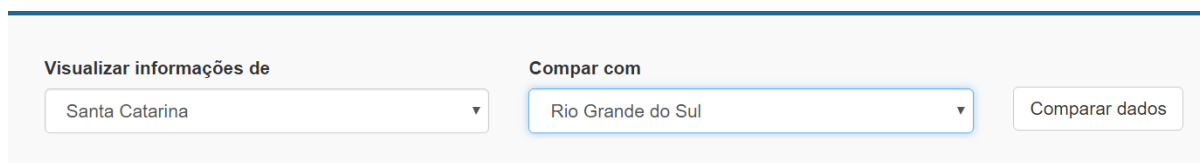
**Figura 24 - Seção comparação por estados.**



Fonte: Autores.

O usuário pode selecionar os estados através de um menu do tipo lista apresentada na Figura 25, para especificar a comparação desejada. Além da comparação entre estados, há também, como já mencionada, a possibilidade de comparar um estado com o Brasil, a fim de verificar o seu desempenho no exame do Enem em relação ao país.

**Figura 25 - Comparação entre estados definida pelo usuário.**



Visualizar informações de

Santa Catarina ▼

Comparar com

Rio Grande do Sul ▼

Comparar dados

**Fonte: Autores**

#### **4.2.3 Visualização por tipos de escola**

A visualização por tipos de escola, conforme a Figura 26, apresenta no eixo vertical a média de notas dos estados em relação à média de notas do Brasil, e no eixo horizontal representa a idade média dos tipos de escola. O objetivo é analisar comparativamente o aproveitamento dos tipos de escola presentes nos Estados. Podendo ainda, visualizar a idade média dos alunos, a fim de descobrir qual faixa etária mais se destacou.

**Figura 26 - Visualização por tipos de escolas.**



**Fonte: Autores.**

A interação nesta sessão, ocorre nas opções de tipos de escolas que o usuário pode selecionar. A Figura 27 apresenta esta interação, possibilitando ao usuário realizar o filtro da informação, com o objetivo de projetá-la de forma mais clara, para que se visualize apenas as informações necessárias.

**Figura 27 - Seleção dos tipos de escolas.**



**Fonte: Autores.**

A área das bolhas representa, proporcionalmente, o número de candidatos que realizaram a prova para cada escola. Quanto maior o número de candidatos, maior será a área.

#### 4.2.4 Visualização por Região

Semelhante à visualização por tipo de Escola, a visualização por região, (Figura 28) tem como objetivo analisar o aproveitamento das escolas públicas por região, através de



comparações. O eixo vertical representa a média de notas dos estados em relação à média de notas do Brasil, e o eixo horizontal representa a idade média dos tipos de escola.

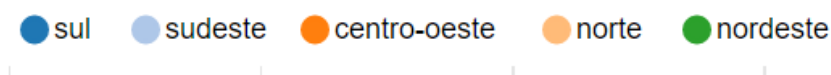
**Figura 28 - Visualização por região.**



**Fonte: Autores.**

O usuário pode selecionar, através das opções, quais regiões serão projetadas pelo algoritmo de visualização. A Figura 29 apresenta este tipo de interação, a qual permite realizar o filtro da informação, e assim, projetar apenas as informações que serão analisadas.

**Figura 29 - Seleção das regiões projetadas.**

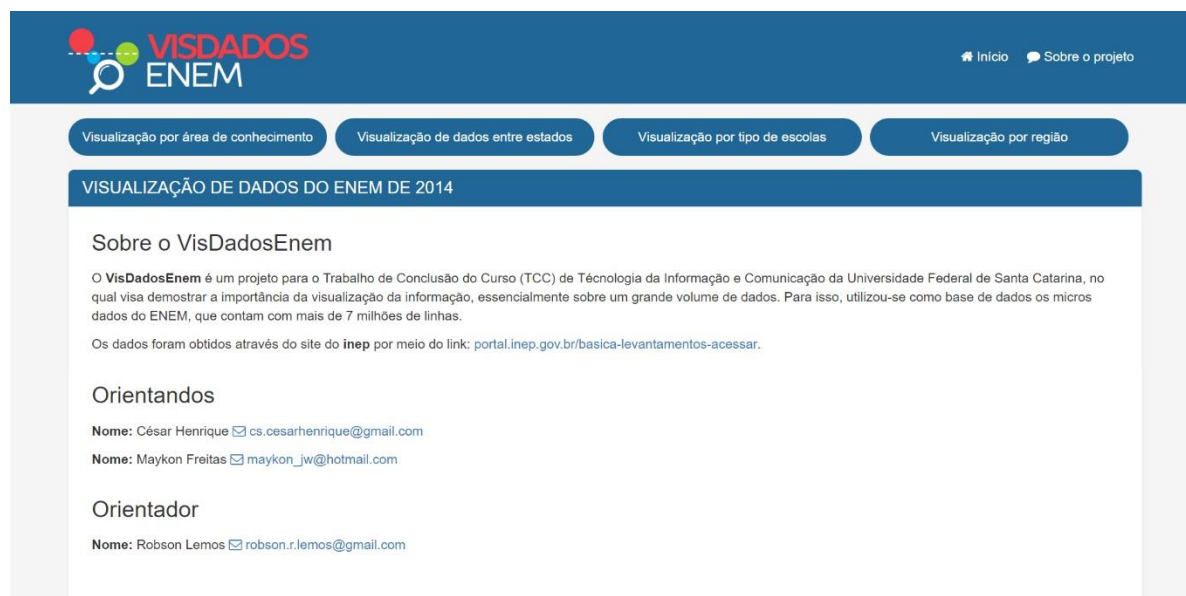


**Fonte: Autores.**

### 4.2.5 Informações Gerais

A Seção de Informações Gerais, com sua interface apresentada na Figura 30, tem a função de esclarecer informações sobre projeto intitulado como VisDadosEnem.

Figura 30 - Seção informações gerais.



Fonte: Autores.

Esta seção traz informações como a descrição do projeto, os autores, o orientador, instituição de origem dos alunos (UFSC) e também o nome do laboratório (LabData - Laboratório de Ciência e Visualização de Dados) que foi um parceiro para a conclusão do trabalho.

## **5 DESENVOLVIMENTO DA APLICAÇÃO WEB**

Este capítulo tem o objetivo de detalhar o processo de desenvolvimento da aplicação Web, bem como métodos de desenvolvimentos, preparação dos dados e desenvolvimento do código da aplicação.

### **5.1 Caracterização dos Dados**

Os dados referentes ao ENEM foram obtidos diretamente no Portal Brasileiro de dados abertos (Dados Abertos, 2016). O portal é uma ferramenta disponibilizada pelo governo com a função de apresentar dados e informações públicas.

A caracterização de dados serve como base para a escolha de uma técnica de visualização ideal, que cumpra o propósito de ampliar o conhecimento utilizando os dois princípios: comunicação e análise.

Sobre os dados do Enem, pode-se identificar uma mistura de dados quantitativos e qualitativos que, respectivamente, utilizam números para demonstrar resultados e categorias para classificar objetos. Ou seja, os atributos, podem assumir valores reais (para as notas obtidas no exame e a quantidade de participantes) e também, valores alfanuméricos (para classificação do tipo de escola e a unidade federativa).

A dimensão apresentada característica do conjunto de dados é definida como bidimensional, que são dados obtidos de pares de variáveis, ou seja, linha e coluna.

### **5.2 Exploração das Técnicas de Visualização**

A técnica utilizada mais adequada para representação visual dos resultados do ENEM foi a Combinação de Técnicas sobre dados Multivariados, que geralmente não apresentam um atributo espacial explícito. Nesta combinação é possível se incluir as seguintes técnicas:

- Técnicas Baseadas em Pontos;
- Técnicas Baseadas em Linhas;

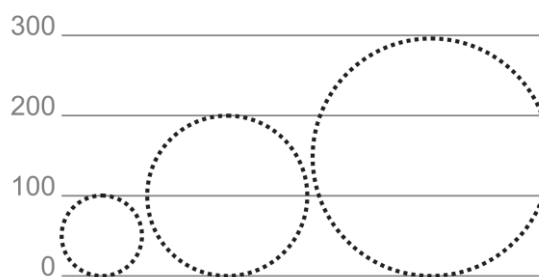
- Técnicas Baseadas em Regiões.

Optou-se por construir uma visualização simples com uma fácil compreensão. Para isso, surgiu a ideia da utilização de bolhas (círculos) para a representação visual. Este tipo de visualização é conhecido como mapa de bolhas (do inglês *Bubble Map*) O círculo é uma figura geométrica plana popularmente conhecida. Com seu design limpo, é caracterizado por estar ligado à definição de circunferência.

De acordo com as variáveis a serem exibidas, foram adicionadas três propriedades no objeto:

- **Área:** Está associada diretamente ao valor numérico da variável. Seu tamanho é proporcional ao valor associado, como se pode observar na Figura 31.

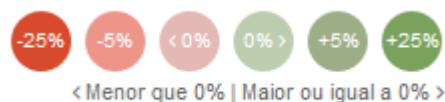
**Figura 31 - Propriedade área.**



**Fonte: Autores.**

- **Cor:** Serve para classificação da variável por intervalos (Figura 32). Ela pode assumir duas cores características, onde uma cor representa o menor valor assumido e a outra, o maior. O intervalo entre estes valores é representado pela variação da intensidade de cada cor. Quanto maior for a intensidade, mais próximo das extremidades será o valor.

**Figura 32 - Legenda para as medias das notas.**



**Fonte: Autores.**

- **Cor da Borda** – Classifica o tipo que a variável pode assumir (Figura 33). Sua cor pode variar de acordo com as tipificações existentes, porém não pode se repetir em um mesmo conjunto, devido a sua unicidade.

**Figura 33 - Legenda da cor da borda para identificar o tipo da variável.**



**Fonte: Autores.**

### 5.3 Desenvolvimento das Funcionalidades do Servidor

Na realização do projeto de visualização de dados, utilizou-se a linguagem de programação PHP que é responsável pelo *back-end* da aplicação, na qual tem como objetivo realizar a população, tratamento e consulta dos dados.

#### 5.3.1 População dos dados

Primeiramente, houve a necessidade de criar um script em PHP para a população dos dados, que foram coletados no Portal Brasileiro de dados abertos ([dados.gov.br](http://dados.gov.br)), no qual é disponibilizado micro dados do ENEM, ENADE, CENSO ESCOLAR, dentre outros. O micro dados do ENEM é disponibilizado em arquivo CSV, que contém aproximadamente 7Gb.

Dessa forma, tornando inviável a edição desse arquivo por editores convencionais, como notepad++, netBeans, dentre outros.

O script em PHP como mostra a Figura 34, realiza a leitura do arquivo CSV obtendo cada linha e transformando a mesma em um *array*. Em seguida faz um tratamento nos campos "NOTA\_CN", "NOTA\_LC", "NOTA\_LC", "NOTA\_MT", "NOTA\_CH", "NU\_NOTA\_REDACAO", este tratamento foi necessário, pois para alguns registros, estes campos não possuíam dado então, caso isso ocorra, o *script* atribui o valor inteiro 0 (zero). Após isto, é realizada a inserção dos dados no banco de dados MYSQL.

**Figura 34 - Script em PHP para importação dos dados do arquivo CSV.**

```

12 $delimitador = ',';
13 $cerca = '';
14 $f = fopen('MICRODADOS_ENEM_2014.csv', 'r');
15 if ($f) {
16     $cabecalho = fgetcsv($f, 0, $delimitador, $cerca);
17     $erros = 0;
18     $msg = "";
19     while (!feof($f)) {
20         $linha = fgetcsv($f, 0, $delimitador, $cerca);
21         if (!$linha) {
22             continue;
23         }
24         $registro = array_combine($cabecalho, $linha);
25         if ($registro['NOTA_CN'] == '') {
26             $registro['NOTA_CN'] = 0;
27         }
28         if ($registro['NOTA_LC'] == '') {
29             $registro['NOTA_LC'] = 0;
30         }
31         if ($registro['NOTA_LC'] == '') {
32             $registro['NOTA_LC'] = 0;
33         }
34         if ($registro['NOTA_MT'] == '') {
35             $registro['NOTA_MT'] = 0;
36         }
37         if ($registro['NOTA_CH'] == '') {
38             $registro['NOTA_CH'] = 0;
39         }
40         if ($registro['NU_NOTA_REDACAO'] == '') {
41             $registro['NU_NOTA_REDACAO'] = 0;
42         }
43
44         $sql = "INSERT INTO enem_2014 (NU_INSCRICAO,UF_RESIDENCIA,IDADE,NOTA_CN,NOTA_CH,"
45             . "NOTA_LC,NOTA_MT,NU_NOTA_REDACAO,Q035) VALUES ('" . $registro["NU_INSCRICAO"]
46             . "','" . $registro["UF_RESIDENCIA"] . "','" . $registro["IDADE"] . "','"
47             . $registro["NOTA_CN"] . "','" . $registro["NOTA_CH"] . "','" . $registro["NOTA_LC"]
48             . "','" . $registro["NOTA_MT"] . "','"
49             . $registro["NU_NOTA_REDACAO"] . "','" . $registro["Q035"] . "')";
50
51         if (mysqli_query($conn, $sql)) {
52         } else {
53             echo "Error: " . $sql . "<br>" . mysqli_error($conn);
54         }
55     }
56     fclose($f);
57 }
58 mysqli_close($conn);

```

**Fonte: Autores.**

### 5.3.2 Desenvolvimento das Classes utilizando PHP

Com o banco de dados inicializado, o próximo passo foi construir uma estrutura de classes em PHP para organizar e dinamizar a obtenção desses dados, conforme o usuário solicite através da aplicação. Optou-se por não utilizar nenhum *framework* PHP sendo que a aplicação tem como objetivo apenas apresentar gráficos de forma interativa, portanto, um *framework*, apenas para isto, acarretaria em um grande volume de código, desnecessário.

Foram elaboradas, ao todo, três classes, que são: Conexão, Dados e Controle.

#### 5.3.3 Classe Conexão

A classe Conexão (código apresentado na Figura 35) é a responsável por efetuar a conexão com o banco de dados MySQL. Através desta classe é possível obter os dados no banco e realizar a sua conversão para um *array*, facilitando, dessa forma, a manipulação dos mesmos.

Figura 35 - Código da classe Conexão.

```

3  class Conexao {
4      public static function getConnection() {
5          return new mysqli('localhost', 'root', '', 'tcc');
6      }
7      public static function fetchArray($query) {
8          $con = self::getConnection();
9          $result = $con->query($query);
10         $row = $result->fetch_array(MYSQL_ASSOC);
11         $con->close();
12         return $row;
13     }
14     public static function transformFetchToArray($query) {
15         $con = self::getConnection();
16         $result = $con->query($query);
17         $cs = self::transformToArray($result);
18         $con->close();
19         return $cs;
20     }
21     private static function transformToArray($consulta) {
22         $arr = array();
23         while ($val = $consulta->fetch_array(MYSQLI_BOTH)) {
24             array_push($arr, $val);
25         }
26         return $arr;
27     }
28     public static function query($query) {
29         $con = self::getConnection();
30         $con->query($query);
31         $id = $con->affected_rows();
32         $con->close();
33         return $id;
34     }
35     public static function create($vars, $tabela) {
36         $con = self::getConnection();
37         $con->createSQL($vars, $tabela);
38         $id = $con->getId();
39         $con->close();
40         return $id;
41     }
42 }

```

Fonte: Autores.

### 5.3.3.1 Classe Dados

Na classe de Dados, encontram-se os métodos responsáveis pela criação dos comandos MSQL e manipulação dos dados obtidos do banco de dados. A Figura 36 mostra a implementação desta classe.



Figura 36. Código da classe Dados.

```

3 class Dados {
4     // Consulta da implementação 1 e 2
5     // - Visualização por área de conhecimento, Visualização de dados entre estados
6     public static function BuscaDadosPorArea() {
7         $query = "SELECT uf_residencia, q035, COUNT(NU_INSCRICAO) as inscricoes,"
8             . " AVG(NOTA_CN) as NOTA_CN, AVG(NOTA_CH) as NOTA_CH, AVG(NOTA_LC) "
9             . " as NOTA_LC, AVG(NOTA_MT) as NOTA_MT, AVG(NU_NOTA_REDACAO) as "
10            . " NU_NOTA_REDACAO "
11            . "FROM enem_2014 WHERE ('A' in (`q035`) OR 'B' in (`q035`) "
12            . "OR 'C' in (`q035`) OR 'D' in (`q035`))"
13            . "group by uf_residencia, q035 "
14            . "order by 1, 2";
15        $estados = Conexao::transformFetchToArray($query);
16        $query = "SELECT q035, AVG(NOTA_CN) as NOTA_CN, AVG(NOTA_CH) as "
17            . "NOTA_CH, AVG(NOTA_LC) as NOTA_LC, AVG(NOTA_MT) as NOTA_MT, "
18            . "AVG(NU_NOTA_REDACAO) as NU_NOTA_REDACAO "
19            . "FROM enem_2014 WHERE ('A' in (`q035`) OR 'B' in (`q035`) "
20            . "OR 'C' in (`q035`) OR 'D' in (`q035`))"
21            . "group by q035 "
22            . "order by 1";
23        $mediasBr = Conexao::transformFetchToArray($query);
24
25        $dados = array();
26        $dados['estados'] = $estados;
27
28        foreach ($mediasBr as $item) {
29            $dados['brasil'][$item['q035']] = $item; //organiza o array com tipo de escola como index
30        }
31
32        return $dados;
33    }
34 }

```

Fonte: Autores.

A função `BuscaDadosPorArea` (descrita na Figura 36) monta os comandos SQL para obter os dados necessários para a visualização das informações para a implementação da sessão Visualização por área de conhecimento e, para implementação da sessão Visualização de dados entre estados. A função `BuscaDadosPorArea` busca o número de inscrições e a média de notas de todos os estados e retorna um *array* com os dados agrupados por estado e tipo de escola, cursada no ensino médio (por exemplo, escola pública, maior parte escola pública, escola particular e maior parte escola particular).

### 5.3.3.2 Classe Controle

A classe Controle, apresentada na Figura 37, tem como finalidade permitir a navegação do usuário na aplicação, através do método GET e também, é responsável pela

criação dos arquivos JSON na qual é realizada a transferência de informações para o cliente. No seguinte trecho de código, pode-se verificar a implementação da classe.

**Figura 37 - Código da classe Controle.**

```

3  class Controle {
4      public static function page() {
5          if (isset($_GET['action'])) {
6              $page = $_GET['action'];
7              switch ($page) {
8                  case 'geral-por-area':
9                      case '': self::geralPorArea();break;
10                     case 'comparacao':self::comparacao();break;
11                     case 'geral-por-escola':self::geralPorEscola();break;
12                     case 'geral-por-regiao':self::geralPorRegiao();break;
13                     case 'sobre':self::sobre();break;
14                     default:self::paginaNaoEncontrada();break;
15                 }
16             } else {
17                 self::geralPorArea();
18             }
19         }
20         //implementação 1 - Visualização por área de conhecimento
21         public static function geralPorArea() {
22             if (!file_exists('arquivos/geral-por-area.json')) {
23                 $arr = Dados::BuscaDadosPorArea();
24                 self::criaJsonGrealPorArea($arr);
25             }
26             include_once 'header.php';
27             include_once 'geral-por-area.php';
28             include_once 'footer.php';
29         }
30     }

```

**Fonte: Autores.**

A função “page” é responsável por controlar a navegação do site, isto é feito, através da leitura do conteúdo, que está na variável \$page, obtida pelo método GET e em seguida, chama a função responsável por exibir o conteúdo da página solicitada.

A função “geralPorArea” é responsável por exibir o conteúdo da sessão Visualização por área de conhecimento, isto é feito, através do comando “include\_once”, onde é chamado os arquivos que contém os conteúdos em HTML. Para que a visualização dos dados seja possível nesta implementação, é preciso que o *front-end* tenha acesso aos dados obtidos com a classe “Dados”. Para tal, criou-se a função “criaJsonGrealPorArea” (descrita na Figura 38), na função em questão, é realizado a organização dos dados obtidos para a criação de um arquivo “.json”, arquivo este que permite a transferência/intercâmbio de dados para o JavaScript.



Figura 38 - Código da função "criaJsonGeralPorArea".

```

32 public static function criaJsonGeralPorArea($arr) {
33     $dadosEstados = $arr['estados'];
34     $dadosBrasil = $arr['brasil'];
35     $estados = array();
36     $escolas = array();
37
38     foreach ($dadosEstados as $item) {
39         $escolas[$item['q035']] = $item;
40         //CALCULO DAS MEDIAS POR AREA DE CONHECIMENTO DO ESTADO COM O BRASIL
41         $escolas[$item['q035']]['COMPARACAO_CN'] = (($item['NOTA_CN'] / $dadosBrasil[$item['q035']]['NOTA_CN']) - 1) * 100;
42         $escolas[$item['q035']]['COMPARACAO_CH'] = (($item['NOTA_CH'] / $dadosBrasil[$item['q035']]['NOTA_CH']) - 1) * 100;
43         $escolas[$item['q035']]['COMPARACAO_LC'] = (($item['NOTA_LC'] / $dadosBrasil[$item['q035']]['NOTA_LC']) - 1) * 100;
44         $escolas[$item['q035']]['COMPARACAO_MT'] = (($item['NOTA_MT'] / $dadosBrasil[$item['q035']]['NOTA_MT']) - 1) * 100;
45         $escolas[$item['q035']]['COMPARACAO_REDACAO'] = (($item['NU_NOTA_REDACAO'] / $dadosBrasil[$item['q035']]['NU_NOTA_REDACAO']) - 1) * 100;
46         //FIM CALCULO DAS MEDIAS
47         $estados[$item['uf_residencia']] = $escolas;
48     }
49     // cria o arquivo .json
50     $fp = fopen('arquivos/geral-por-area.json', "w");
51     fwrite($fp, json_encode($estados));
52     fclose($fp);
53 }
54

```

Fonte: Autores.

## 5.4 Desenvolvimento de Funcionalidades para o Cliente

No desenvolvimento do *front-end* foi utilizado o HTML5 juntamente com o *framework* Bootstrap para CSS3 e as bibliotecas jQuery e D3 para JavaScript. Para a utilização dos mesmos é necessária a inclusão dos respectivos arquivos dentro da tag <head> no HTML como apresentado na Figura 39.

Figura 39 - Referencias no HTML para os *frameworks* utilizados.

```

1 <!doctype html>
2 <html lang="pt-br">
3
4 <head>
5     <meta charset="ISO-8859-1">
6     <title>VisDadosEnemD3</title>
7     <link rel="stylesheet" href="assets/css/bootstrap.css" type="text/css" />
8     <script src="assets/js/jquery-3.1.1.min.js" type="text/javascript"></script>
9     <script src="assets/js/d3.v3.min.js"></script>
10 </head>

```

Fonte: Autores

Na implementação da funcionalidade que realiza comparações de desempenho entre estados foi utilizado o elemento do HTML <select> para realizar a interação do usuário com a aplicação, conforme se observa na Figura 40.

**Figura 40 - Código HTML utilizando o elemento <select>.**

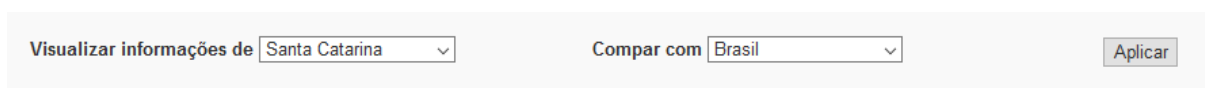
```

7  <div class="col-md-12 selects">
8
9  <div class="left-col clearfix">
10
11    <div class="col-md-6 text-center">
12      <label for="uf">Visualização de dados de</label>
13      <select name="uf" id="uf" class="select-style">
14        <option selected="selected" value="SC">Santa Catarina</option>
15      </select>
16    </div>
17    <div class="col-md-6 text-center">
18      <label for="comprar" >comparado com</label>
19      <select name="comprar" id="comprar" class="select-style">
20        <option value="TO">Tocantins</option>
  
```

**Fonte: Autores.**

A Figura 41 mostra a visualização no navegador sem estilização de CSS.

**Figura 41 - Navegador sem estilização do CSS.**



Visualizar informações de  Comparar com

**Fonte: Autores.**

Para que os seletores e o botão ficassem mais intuitivos e visualmente, com aparência mais adequada ao *layout*, foram criadas as linhas de CSS, apresentadas na Figura 42.

Figura 42 - Folha de estilo para deixar os seletores e o botão mais intuitivos e visualmente.

```

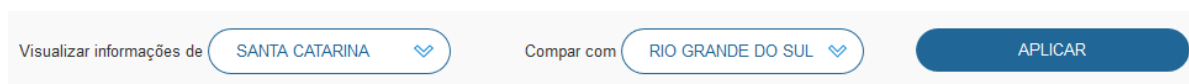
260  .btn-style {
261      background: #206696;
262      border-radius: 43px;
263      color: white ;
264      padding: 10px 20px;
265      text-transform: uppercase;
266      transition: ease-in-out 0.3s;
267      border: 0 none;
268  }
269  .btn-style:hover{
270      background: #00AFEF;
271      color: #fff;
272  }
273  .select-style {
274      background: white none repeat scroll 0 0;
275      border: 1px solid #206696;
276      border-radius: 43px;
277      color: #206696;
278      padding: 10px 20px;
279      text-transform: uppercase;
280      cursor: pointer;
281      -moz-appearance:none;
282      -webkit-appearance: none;
283      background-image: url('../img/down-arrow.png');
284      background-position: calc(100% - 15px) center ;
285      background-repeat: no-repeat;
286      background-size: 17px;
287
288      transition: ease-in-out 0.3s;
289  }
290  .select-style:hover, .select-style:focus, .select-style:active {
291      border-color: #00AFEF;
292      background-color: #00AFEF;
293      color: white;
294  }

```

Fonte: Autores.

O resultado da folha de estilo descrita anteriormente, pode ser visualizado na Figura 43.

Figura 43 - Navegação com estilização do CSS.



Fonte: Autores.

Na Figura 44, pode-se observar o código em JavaScript responsável pela recuperação dos dados informados pelo usuário para a implementação da sessão Visualização por área de conhecimento.

**Figura 44 - Código para recuperação dos dados informados pelo usuário.**

```

8  $( '.checkboxes input' ).change( function () {
9      $( '.box-svg > div' ).remove();
10     geral();
11 });
12
13 function geral() {
14     var cn = $( '#cb1' ).is( ':checked' );
15     var ch = $( '#cb2' ).is( ':checked' );
16     var lc = $( '#cb3' ).is( ':checked' );
17     var mt = $( '#cb4' ).is( ':checked' );
18     var re = $( '#cb5' ).is( ':checked' );
19     if ( cn == '' && ch == '' && lc == '' && mt == '' && re == '' ) {
20         alert( 'Selecione ao menos uma área de conhecimento' );
21         return 0;
22     }

```

**Fonte: Autores.**

A recuperação dos dados informados pelo usuário é realizada quando o usuário clica no *checkbox*, o javascript identifica a ação do usuário, por meio do método `change()` com jQuery e chama a função `geral()`. Na função “geral” é identificado quais áreas estão marcadas (Figura 45) através do método “`is(‘checked’)`”

**Figura 45 – Seleções por área de conhecimento.**

Selecione a(s) área(s) de conhecimento:

- ☐ Ciências da Natureza
- ☐ Ciências Humanas
- ☐ Linguagens e Códigos
- ☐ Matemática
- ☒ Redação

**Fonte: Autores.**

Após a identificação das áreas, faz-se a leitura do arquivo “geral-por-area.json”, que contém os dados agrupados por estado e tipo de escola e realiza a comparação das médias de notas das áreas selecionadas de cada estado com o Brasil. Em seguida, envia-se as informações como parâmetro para a função “D3start”, função que é responsável por converter

os dados em gráficos para a implementação da sessão Visualização por área de conhecimento e da sessão Visualização de dados entre estados, conforme a Figura 46.

**Figura 46 - Função que converte dados em gráficos.**

```

194 d3.json('arquivos/geral-por-area.json', function (myData) {
195   var calcMedia = new Array();
196   calcMedia['A'] = 0, calcMedia['B'] = 0, calcMedia['C'] = 0, calcMedia['D'] = 0;
197   var i1 = 0;
198   $.each(myData, function (index, value) { //PERCORRE CADA ESTADO
199     var ufAtual = new Array(); var i2 = 0;
200     $.each(value, function (index2, value2) { //PERCORRE CADA ESCOLA
201       var mediaEsc = 0.00;
202       var mediaComp = 0.00;
203       var nAreas = 0;
204       if (cn) {
205         mediaComp = mediaComp + parseFloat(value2.COMPARACAO_CN);
206         mediaEsc = mediaEsc + parseFloat(value2.NOTA_CN);
207         nAreas++;
208       }
209       if (ch) {
210         mediaComp = mediaComp + parseFloat(value2.COMPARACAO_CH);
211         mediaEsc = mediaEsc + parseFloat(value2.NOTA_CH);
212         nAreas++;
213       }
214       if (lc) {
215         mediaComp = mediaComp + parseFloat(value2.COMPARACAO_LC);
216         mediaEsc = mediaEsc + parseFloat(value2.NOTA_LC);
217         nAreas++;
218       }
219       if (mt) {
220         mediaComp = mediaComp + parseFloat(value2.COMPARACAO_MT);
221         mediaEsc = mediaEsc + parseFloat(value2.NOTA_MT);
222         nAreas++;
223       }
224       if (re) {
225         mediaComp = mediaComp + parseFloat(value2.COMPARACAO_REDACAO);
226         mediaEsc = mediaEsc + parseFloat(value2.NU_NOTA_REDACAO);
227         nAreas++;
228       }
229       mediaComp = mediaComp / nAreas;
230       mediaEsc = mediaEsc / nAreas;
231       indexCor = getIndiceCor(mediaComp);
232       var aux = {"cor": indexCor, "tipo": i2, "codigo_cor": indexCor,
233         "inscritos": value2.inscricoes, "media": mediaEsc, "relacao": mediaComp,
234         "pays": "a", "pop": value2.inscricoes, "uf": value2.uf_residencia};
235       ufAtual.push(aux);
236       if (i2 >= 3) { i2 = 0; } else { i2++; }
237     });
238     var canvasW = 274; var canvasH = 280; var divisorEscala = 5; var maxRadius = 56;
239     D3start(ufAtual, index, 'estado2', i1, canvasW, canvasH, divisorEscala, maxRadius);
240     i1++;
241   });
242 });
243

```

**Fonte: Autores.**

A biblioteca d3.js foi utilizada para realizar a conversão dos dados em formas gráficas para a visualização da informação. Na figura 47, pode-se observar a criação de um elemento svg com a forma de um círculo.



**Figura 47 - Criação de um Círculo em svg utilizando o D3.**

```
1  var canvas = d3.select(".box-svg")
2      .append("svg")
3      .attr("width", 500)
4      .attr("height", 500);
5
6  var circleA = canvas
7      .append("circle")
8      .attr("cx", 100)
9      .attr("cy", 150)
10     .attr("r", 50);
11
```

**Fonte: Autores.**

O código representado na Figura 47, resulta no elemento apresentado pela Figura 48.

**Figura 48 - Circulo criado com a D3.**



**Fonte: Autores.**

Para que o círculo indique informações, utilizou-se alguns atributos como o diâmetro, cor de preenchimento e cor da borda. No trecho de código apresentado pela Figura 49 é possível visualizar a adição desses atributos

Figura 49 - Criando círculos com atributos.

```

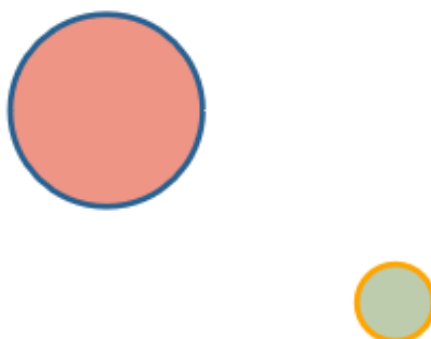
17     .append("svg")
18     .attr("width", 500)
19     .attr("height", 500);
20
21     var circleA = canvas
22     .append("circle")
23     .attr("cx", 100)
24     .attr("cy", 150)
25     .attr("r", 50)
26     .attr("fill", '#EE9586')
27     .attr("stroke", '#286090')
28     .attr("stroke-width", 3);
29
30     var circleB = canvas.
31     .append("circle")
32     .attr("cx", 250)
33     .attr("cy", 250)
34     .attr("r", 20)
35     .attr("fill", '#BECCAE')
36     .attr("stroke", '#FFA500')
37     .attr("stroke-width", 3);

```

Fonte: Autores.

O resultado da implantação (Figura 49) pode ser visto na Figura 50.

Figura 50 - Resultado da implementação dos círculos com atributos.



Fonte: Autores.

A cor da borda representa o tipo de escola (por exemplo, pública, maior parte pública, particular e maior parte particular), a cor de preenchimento representa a comparação da média de notas com um estado ou o país e o diâmetro indica o número de inscritos no exame.

Utilizou-se também de alguns métodos de animações do d3 para que os círculos não ficassem dispostos de maneira estática, dessa forma transmitindo uma maior impressão de interatividade como pode ser observado na Figura 51.

**Figura 51 - Implementação de um dos métodos de animação do d3.**

```

41 function drawBubbles(arrUf) {
42     nodes = convertData(arrUf);
43     var force = d3.layout.force()
44         .nodes(nodes)
45         .size([width, height])
46         .gravity(.02)
47         .charge(0)
48         .on("tick", tick)
49         .start();
50
51     node.transition()
52         .duration(750)
53         .delay(function(d, i) {
54             return i * 5;
55         })
56         .attrTween("r", function(d) {
57             var i = d3.interpolate(0, d.radius);
58             return function(t) {
59                 return d.radius = i(t);
60             };
61         });

```

**Fonte: Autores.**

Na implementação da sessão Visualização por tipos de escolas e da sessão Visualização por região, utilizou-se a biblioteca NVD3 por disponibilizar gráficos complexos prontos, utilizando o D3 como base. Para a criação do gráfico da implementação da sessão Visualização por tipos de escolas, utilizou-se do seguinte script (Figura 52).

**Figura 52 - Criação do gráfico da sessão Visualização por tipos de escolas.**

```

74 nv.addGraph(function () {
75   d3.json('arquivos/geral-por-escola.json', function (myData) {
76     var chart = nv.models.scatterChart()
77       .showDistX(true)
78       .showDistY(true)
79       .transitionDuration(350);
80     chart.x(function (d) {
81       chart.tooltipContent(function (key) {
82         var tooltipString = '';
83         tooltipString += '<div class="tooltip-3">';
84         tooltipString += '<h3 class="title">' + getNomeEstado(d.uf) + '</h3>';
85         tooltipString += '<p><b>Oriundos ' + key + '</b></p>';
86         tooltipString += '<p><b>Inscritos: </b>' + d.inscritos + '</p>';
87         tooltipString += '<p><b>Média de notas do estado: </b>' + d.media + '</p>';
88         tooltipString += '<p><b>Comparação de média de notas com o Brasil: </b>' + d.comparacao + '%</p>';
89         tooltipString += '</div>';
90         return tooltipString;
91       });
92       return d.x;
93     });
94     chart.xAxis.tickFormat(d3.format('.02f'));
95     chart.yAxis.tickFormat(d3.format('.02f'));
96     chart.scatter.onlyCircles(true);
97     d3.select('#chart svg')
98       .datum(myData)
99       .call(chart);
100     nv.utils.windowResize(chart.update);
101     return chart;
102   });
103 });

```

**Fonte; Autores.**

Para o código interpretar os dados criou-se um arquivo JSON com os dados organizados da forma como é apresentado na Figura 53.

**Figura 53 - Código em JSON para interpretar os dados.**

```

1  [{
2    "key": "escola publica",
3    "values": [{
4      "shape": "circle",
5      "size": 4.1232,
6      "m": "292.57959837",
7      "x": "27.24107489328677",
8      "y": "-2.2596685036839",
9      "uf": "AC",
10     "media": "292,58",
11     "inscritos": "41.232,00",
12     "comparacao": "-2,26",
13     "series": 0
14   }],

```

**Fonte: Autores.**

Na implementação aproveitou-se o mesmo código JavaScript da terceira implementação, alterando apenas a organização dos dados no arquivo JSON, conforme a Figura 54.

**Figura 54 - Código alterado para ser utilizado na terceira implementação.**

```

1      [{
2
3      -      "key": "sul",
4      -      "values": [{
5      -          "shape": "circle",
6      -          "size": 26.6092,
7      -          "m": "311.71256257",
8      -          "x": "24.713332230957715",
9      -          "y": 4.1319673924398,
10     -          "uf": "PR",
11     -          "media": "311,71",
12     -          "inscritos": "266.092,00",
13     -          "comparacao": "4,13",
14     -          "series": 0
15     -      },

```

**Fonte: Autores.**

Após a apresentação do desenvolvimento da aplicação, parte-se para a apresentação dos resultados obtidos com este trabalho.

## 6 RESULTADOS

Pode-se dizer que, de modo geral, este trabalho compriu com seu objetivo geral: projetar e desenvolver uma ferramenta para visualização de dados. A biblioteca D3 auxiliou diretamente na implementação das técnicas de visualização e tornou possível a criação de quatro funcionalidades para a análise comparativa dos dados do ENEM.

Qualquer aplicação ou site, que necessite ser publicado na Web, precisa utilizar um serviço hospedeiro para armazenamento do projeto e um endereço (domínio) para ser acessado. Com o VisdadosEnem não é diferente, como seu conteúdo foi projetado especificamente para a Web, foi necessário fazer uso destes recursos.

Por se tratar de um trabalho para conclusão do curso de Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina (UFSC), foi solicitado o fornecimento dos recursos ao Laboratório, para então, finalizar o projeto.

O laboratório de ciência e visualização de dados (LabData) da UFSC atendeu ao pedido, e solicitou um espaço no servidor da SETIC (Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação) para hospedar todo conteúdo da ferramenta VisdadosEnem, que pode ser acessada pelo domínio <http://labdata.sites.ufsc.br/visdadosenem>.

Levando em consideração os recursos de visualização que o VisdadosEnem proporciona, comprovou-se que seu uso, ou o uso de ferramentas similares, faz-se necessário para realizar análises sobre um grande conjunto de dados, como no caso dos micro-dados do ENEM. Análise esta, que uma tabela comum dificilmente iria proporcionar, devido à grande extensão de dados, o que dificultaria a visualização.

Como resultado das análises comparativas realizadas sobre os dados do ENEM, pode-se citar:

Os alunos de escolas particulares, ou que cursaram maior parte do tempo nestas, obtiveram, no geral, uma melhor média frente às outras escolas (Figura 55).

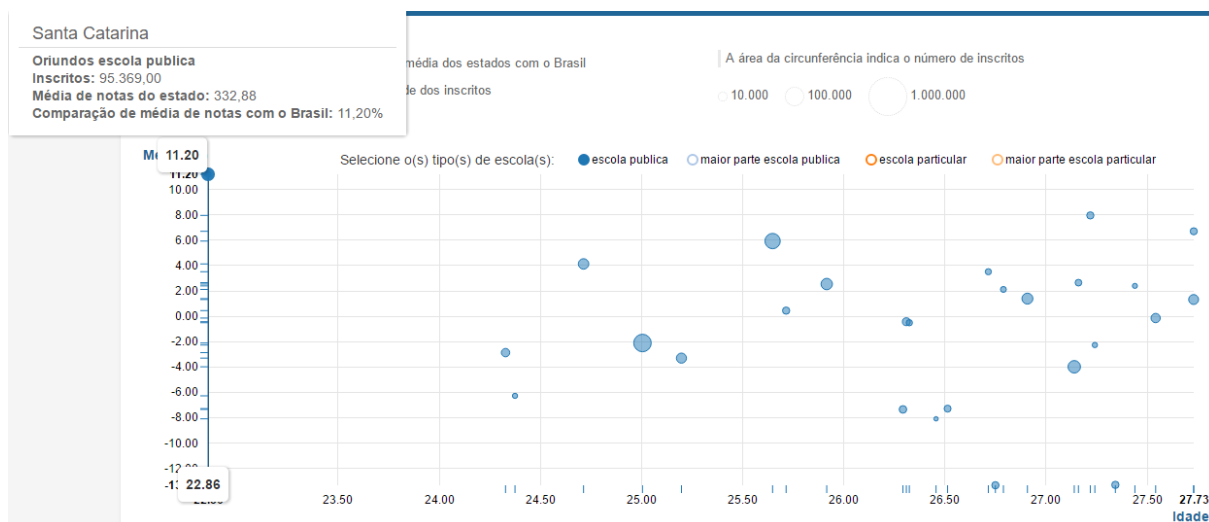
**Figura 55 - Visualização por área de conhecimento**



**Fonte: Autores.**

Os alunos oriundos de escola pública de Santa Catarina possuem a média de idade mais jovem em relação aos demais estados e o estado alcançou a maior média geral do país. (Figura 56).

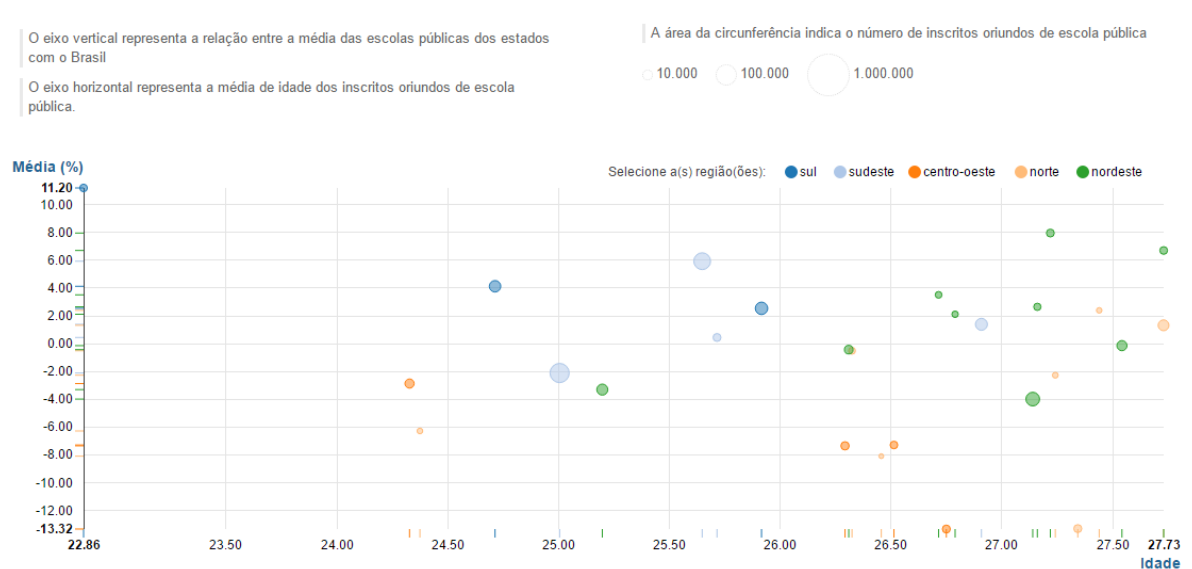
Figura 56 - Visualização por tipo de escola



Fonte: Autores.

A média de escola pública da região sul foi a mais alta em comparação à média geral, enquanto que a região centro-oeste obteve o pior desempenho entre as cinco regiões. (Figura 57).

Figura 57 – Visualização por região



Fonte: Autores.

Tendo sido aprendido os principais resultados do trabalho, parte-se para a seção de considerações finais e trabalhos futuros.



## 7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O aumento em massa dos dados, na área da informática, trouxe grandes problemas para análise e compreensão da informação. Devido ao volume excessivo de dados, fator este que dificulta o processo de visualização, torna-se necessário o uso de ferramentas auxiliares com o objetivo de contribuir para melhoria deste processo.

O estudo das Tecnologias da Informação e Comunicação possibilitou o desenvolvimento da ferramenta para visualização de dados, intitulado VisdadosEnem. O projeto foi desenvolvido utilizando-se um conjunto de dados, os micro dados do ENEM, para comprovar a necessidade e eficácia da utilização de ferramentas de visualização.

Pode-se afirmar que, a utilização do VisdadosEnem possibilitou uma análise comparativa sobre o conjunto dados. Foram descobertos padrões, que seriam dificilmente encontrados em tabela ou gráfico tradicionais.

Além disso, o projeto proporcionou uma ferramenta *on-line* de utilidade pública, que pode ser facilmente acessada e utilizada para diversas análises. A mesma pode ser de interesse aos órgãos educacionais, que podem analisar os resultados dos tipos de escola, estados e regiões e assim, contribuir para a tomada de decisão com base nas informações extraídas com a visualização.

Este projeto inicia várias possibilidades de trabalhos e estudos futuros. O conjunto de dados utilizado para projeção da visualização e as informações que foram obtidas podem ser utilizadas em diferentes cenários. Como exemplo, estudos estatísticos sobre o ENEM, como proceder à avaliação da qualidade da escola pública em comparação à privada em relação às notas obtidas, identificar facilmente quais os estados que possuem um melhor desempenho no ENEM.

Enfim, várias possibilidades de arranjos de dados foram apresentadas. Porém, a utilização de todo o potencial do VisDadosEnem somente poderá ser alcançada, se o usuário souber utilizar essas ferramentas conforme seus objetivos e assim possibilitando um importante apoio à tomada de decisão.

## 8 REFERÊNCIAS

**BOOTSTRAP**. Disponível em: <<http://getbootstrap.com.br/>>. Acesso em: 30 out. 2016.

BORTOLOSSI, Humberto José. **Criando conteúdos educacionais digitais interativos em matemática e estatística com o uso integrado de tecnologias: GeoGebra, JavaView, HTML, CSS, MathML e JavaScript**. Instituto Geogebra, São Paulo, v. 1, n. 1, p. XXXVIII – XXXVI, nov. 2012. Disponível em: <<http://revistas.pucsp.br/index.php/IGISP/issue/view/557/showToc>>. Acesso em: 21 nov. 2016.

Card, K. S., Mackinlay, J. D., and Shneiderman, B. (1999). **Readings in Information Visualization, using vision to think**. Morgan Kaufmann, Cal. USA.

Few, S. **Now you see it: simple visualization techniques for quantitative analysis**, Analytics Press, 2009.

Freitas, C. M. D. S e Wagner, F. R. Fernandes de. **Suporte às tarefas da análise exploratória visual**. Revista de Informática Teórica e Aplicada, 2(1): 5-36, jan. 1995.

GRANNELL, C. (2007). **The Essential Guide to CSS and HTML Web Design**. New York: APress.

**JSON**. Disponível em: <<http://json.org/json-pt.html/>>. Acesso em: 30 out. 2016.

LAMIM, Jonathan. **Afinal, o que é Front-end e o que é Back-end?** 2014. Disponível em: <<https://www.oficinadanet.com.br/post/13541-afinal-o-que-e-front-end-e-o-que-e-Back-end->>. Acesso em: 15 setembro 2016.

LUZZARDI, P. R. G. **CrITÉrios de Avaliação de Técnicas de Visualização de Informações Hierárquicas**. Tese de doutorado, UFRGS, Porto Alegre, 2003.

M. O. Ward, B. N. Lipchak. “**A Visualization Tool for Exploratory Analysis of Cyclic Multivariate Data.**” *Metrika* 51:1 (2000), 27–38.

Matthew O. Ward. “**A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization.**” *Information Visualization* 1:3–4 (2002), 194–210.

NAVARRO, Rhea Mae. **What is HTML?** Disponível em: <<http://ctu-ict.net/wp/index.php/2015/10/14/html-vs-css/>>. Acesso em: 1 setembro 2016.

**NEWSMAP.** Disponível em: <<http://newsmap.jp/>>. Acesso em: 17 jul. 2016.

**NVD3.** Disponível em: <<http://nvd3.org/>>. Acesso em: 30 out. 2016.

Porta de Brasileiro de Dados Abertos. Disponível em: <<http://dados.gov.br/>>. Acesso em: 30 out. 2016.

W3C ESCRITÓRIO BRASIL. Disponível em: <[www.w3c.br](http://www.w3c.br/)>. Acesso em: 14 set. 2016.

Ward, M., Grinstein, G., Kein, D. **Interactive data visualization: foundations, techniques, and applications**, CRC Press, 2015.

Ware, C. **Information Visualization: perception for Design**, San Francisco, Morgan Kaufmann, 2000.

**WORD Clouds.** Disponível em: <<https://www.jasondavies.com/wordcloud/>>. Acesso em: 14 nov. 2016.

**WORM Circuito Explorador.** 2004. Disponível em: <[elegans.herokuapp.com](http://elegans.herokuapp.com)>. Acesso em: 15 nov. 2016.